PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/63, 15/10, 15/81, C12Q 1/68

A2

(11) International Publication Number: WO 98/13502

(43) International Publication Date: 2 April 1998 (02.04.98)

(21) International Application Number: PCT/US97/17276

(22) International Filing Date: 26 September 1997 (26.09.97)

(30) Priority Data:
08/721,730 27 September 1996 (27.09.96) US

(71) Applicant: ICOS CORPORATION [US/US]; 22021 20th Avenue, S.E., Bothell, WA 98021 (US).

(72) Inventors: GOODMAN, Richard, H.; 18560 Westview Drive, Lake Oswego, OR 97034-7382 (US). HOEKSTRA, Merl, F.; 10321 216th, S.E., Snohomish, WA 98290 (US).

(74) Agent: WILLIAMS, Joseph, A., Jr.; Marshall, O'Toole, Gerstein, Murray & Borun, 6300 Sears Tower, 233 S. Wacker Drive, Chicago, IL 60606-6402 (US).

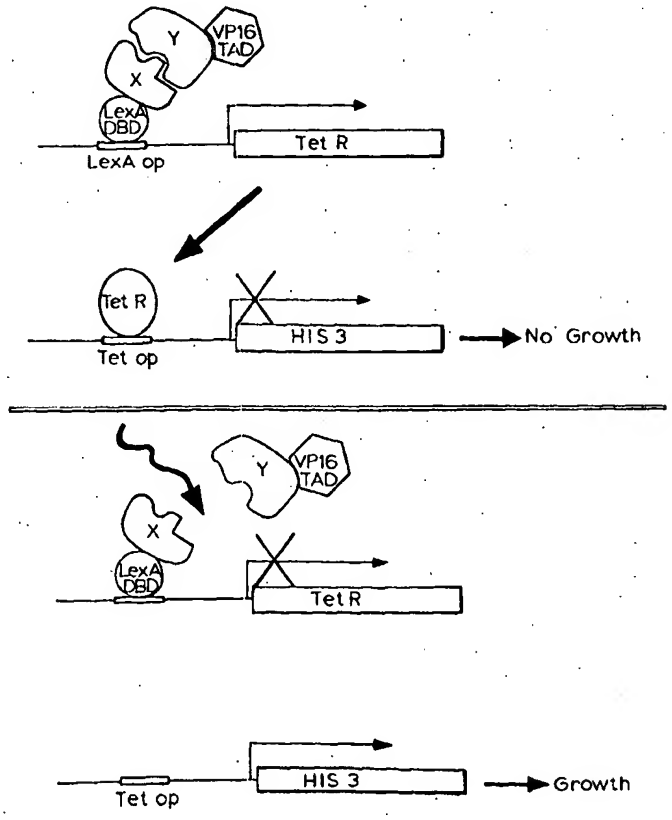
(81) Designated States: AU, BR, CA, CN, CZ, FI, HU, IL, JP, MX, NO, PL, RU, SK, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD TO IDENTIFY COMPOUNDS FOR DISRUPTING PROTEIN/PROTEIN INTERACTIONS

(57) Abstract

The present invention relates generally to materials and methods for identification of inhibitors of interactions between known binding partner proteins.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHODS TO IDENTIFY COMPOUNDS FOR DISRUPTING PROTEIN/PROTEIN INTERACTIONS

Background of the Invention

The present invention relates to a novel method to identify inhibitors of protein/protein interactions.

Background

Modulation of protein/protein interactions is an attractive target for drug discovery and development. Potential methods by which drugs can regulate protein/protein interactions are numerous, including, for example, regulation of expression of one or more of the binding proteins, modulation of post-translational modification, and direct interference with the capacity of one protein to bind to one or more binding partners. More importantly, recent observations make it increasingly clear that supramolecular protein complexes, involving two or more binding proteins, play an important and essential roles in signal transduction, gene expression, cell proliferation and duplication, and cell cycle progression. For example, in the repair of UV damaged DNA, a so-called "repairsome" that contains over ten individual proteins is assembled into a complex which can then carry out the necessary repair. Likewise, gene transcription occurs through the concerted action of greater than twenty proteins. Signal transduction proteins, such as receptor protein kinases, are part of large complexes with many proteins. Contacts through *Src* homology type 2 (SH2) domains on the receptor kinases, for example, are noteworthy protein interaction which are part of one or more enzymatic cascade important for many metabolic processes. Disrupting the binding capacity of one or more proteins which form any of these larger complex is therefore an important and untapped method to control action of the overall complex.

Protein/protein interactions have been discovered and characterized by a variety of methods: (i) standard biochemical affinity

methods such as chromatography or co-immunoprecipitations; (ii) gel overlay methods; (iii) co-purification by traditional biochemistry; and (iv) two-hybrid analysis [Fields and Song, *Nature* 340:245-246 (1989); Fields, *Methods: A Companion to Methods in Enzymology* 5:116-124 (1993); U.S. Patent 5,283, 173 issued February 1, 1994 to Fields, *et al.*]. The most recent of these approaches, the two hybrid method, has enjoyed broad application because of its relative ease of use for gene identification from cDNA fusion libraries. [See Chien *et al., Proc. Natl. Acad. Sci. (USA)* 88:9578-9582 (1991); Dalton and Treisman, *Cell* 72:223-232 (1993); and Durfee, *et al., Genes and Devel.* 7:555-569 (1993)].

The two hybrid system is based on targeting and identifying a protein/protein interaction through the use of a reporter system. The described two hybrid systems either use the yeast Gal4 DNA binding domain or the *E. coli* lexA DNA binding domain and couple this region to a transcriptional activator such as Gal4 or VP16 that drives a reporter like β galactosidase or HIS3.

In principle the two hybrid assay could be used for drug screening. [See WO 96/03501 and WO 96/03499.] In such a scenario, loss of β galactosidase or HIS3 activity would be identified after the yeast strain is treated with a compound. In practice, however, use of the two hybrid system is technically undesirable for several reasons. In instances where the β galactosidase or HIS3 protein are employed as the reporter protein, a loss of activity is particularly difficult to detect because the expressed reporter protein is too long lived to be used in a high throughput mode. If a candidate binding inhibitor compound is metabolized faster than the previously expressed reporter protein is turned over, it is difficult to detect inhibitory action of the candidate drug while a reporter protein is still active. In high throughput screening, the loss of a positive signal, for example, β galactosidase or HIS3 is impossible to detect. Present robotocized screening and detection methods are simply not sufficiently sensitive or robust to detect loss of a signal.

Thus there is a need in the art to develop a rapid screening method that gives a positive signal, as opposed to a negative signal, when a protein/protein interaction is disrupted. Such a system must be capable of using protein interactions that are initially detected by any of the above mentioned approaches and must be sufficiently robust to detect a gain of function when a protein interaction is lost. In essence, the screening method must give a signal when an interaction is lost, not lose a signal when an interaction is lost. Such a system must be sensitive to subtle interactions, in particular ones that are caused by post-translational modification like protein phosphorylation. Finally for large scale screening, such as high throughput screening, the system must be manipulable such that a large signal-to-noise ration can be easily detected.

Brief Summary of the Invention

In one aspect, the present invention provides materials that are useful for the identification of compounds which inhibit interaction between known binding partner proteins. See Figure 1. The invention provides host cells transformed or transfected with DNA comprising: (i) a repressor gene encoding DNA binding protein that acts as a repressor protein, said repressor gene under transcriptional control of a promoter; (ii) a selectable marker gene encoding a selectable marker protein; said selectable marker gene under transcriptional control of an operator; said operator regulated by interaction with said repressor protein; (iii) a first recombinant fusion protein gene encoding a first binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activating protein or a transactivating domain of a transcriptional activating protein; and (iv) a second recombinant fusion protein gene encoding a second binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activating protein or a transactivating domain of a transcriptional activating protein, whichever domain is not encoded by the first

fusion protein gene, said second binding protein or binding fragment thereof capable of interacting with said first binding protein or binding fragment thereof such that interaction of said second binding protein or binding fragment thereof and said first binding protein or binding fragment thereof brings into proximity a DNA binding domain and a transactivating domain forming a functional transcriptional activating protein; said functional transcriptional activating protein acting on said promoter to increase expression of said repressor gene.

The invention comprehends host cells wherein the various genes and regulatory sequences are encoded on a single DNA molecule as well as host cells wherein one or more of the repressor gene, the selectable marker gene, the first recombinant fusion protein gene, and the second recombinant fusion protein gene are encoded on distinct DNA expression constructs. In a preferred embodiment, the host cells are transformed or transfected with DNA encoding the repressor gene, the selectable marker gene, the first recombinant fusion protein gene, and the second recombinant fusion protein gene, each encoded on a distinct expression construct. Regardless of the number of DNA expression constructs introduced, each transformed or transfected DNA expression construct further comprises a selectable marker gene sequence, the expression of which is used to confirm that transfection or transformation was, in fact, accomplished. Selectable marker genes encoded on individually transformed or transfected DNA expression constructs are distinguishable from the selectable marker under transcriptional regulation of the *tet* operator in that expression of the selectable marker gene regulated by the *tet* operator is central to the preferred embodiment; *i.e.*, regulated expression of the selectable marker gene by the *tet* operator provides a measurable phenotypic change in the host cell that is used to identify a binding protein inhibitor. Selectable marker genes encoded on individually transformed or transfected DNA expression constructs are provided as determinants of successful transfection or transformation of the individual

DNA expression constructs. Preferred host cells of the invention include transformed *S. cerevisiae* strains designated YI596 and YI584 which were deposited August 13, 1996 with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland 20852, and assigned Accession Numbers ATCC 74384 and ATCC 74385, respectively.

The host cells of the invention include any cell type capable of expressing the heterologous proteins required as described above and which are capable of being transformed or transfected with functional promoter and operator sequences which regulate expression of the heterologous proteins also as described. In a preferred embodiment, the host cells are of either mammal, insect or yeast origin. Presently, the most preferred host cell is a yeast cell. The preferred yeast cells of the invention can be selected from various strains, including the *S. cerevisiae* yeast transformants described in Table 1. Alternative yeast specimens include *S.pombe, K.lactis, P.pastoris, S.carlsbergensis* and *C.albicans*. Preferred mammalian host cells of the invention include Chinese hamster ovary (CHO), COS, HeLa, 3T3, CV1, LTK, 293T3, Rat1, PC12 or any other transfectable cell line of human or rodent origin. Preferred insect cells lines include SF9 cells.

In a preferred embodiment, the selectable marker gene is regulated by an operator and encodes an enzyme in a pathway for synthesis of a nutritional requirement for said host cell such that expression of said selectable marker protein is required for growth of said host cell on media lacking said nutritional requirement. Thus, as in a preferred embodiment where a repressor protein interacts with the operator, transcription of the selectable marker gene is down-regulated and the host cells are identified by an inability to grow on media lacking the nutritional requirement and an ability to grow on media containing the nutritional requirement. In a most preferred embodiment, the selectable marker gene encodes the HIS3 protein, and host cells transformed or transfected with a HIS3-encoding DNA expression construct are selected following growth on media in the presence

and absence of histidine. The invention, however, comprehends any of a number of alternative selectable marker genes regulated by an operator. Gene alternatives include, for example *URA3, LEU2, LYS2* or those encoding any of the multitude of enzymes required in various pathways for production of a nutritional requirement which can be definitively excluded from the media of growth. In addition, conventional reporter genes such as chloramphenicol acetyltransferase (CAT), firefly luciferase, β-galactosidase (β-gal), secreted alkaline phosphatase (SEAP), green fluorescent protein (GFP), human growth hormone (hGH), β-glucuronidase, neomycin, hygromycin, thymidine kinase (TK) and the like may be utilized in the invention.

In the preferred embodiment, the host cells include a repressor protein gene encoding the tetracycline resistance protein which acts on the *tet* operator to decrease expression of the selectable marker gene. The invention, however, also encompasses alternatives to the *tet* repressor and operator, for example, *E.coli trp* repressor and operator, *his* repressor and operator, and *lac* operon repressor and operator.

The DNA binding domain and transactivating domain components of the fusion protein may be derived from the same transcription factor or from different transcription factors as long as bringing the two domains into proximity permits formation of a functional transcriptional activity protein that increases expression of the repressor protein with high efficiency. A high efficiency transcriptional activating protein is defined as having both a DNA binding domain exhibiting high affinity binding for the recognized promoter sequence and a transactivating domain having high affinity binding for transcriptional machinery proteins required to express repressor gene mRNA. The DNA binding domain component of a fusion protein of the invention can be derived from any of a number of different proteins including, for example, LexA or Gal4. Similarly, the transactivating component of the invention's fusion proteins can be derived from a number of different transcriptional activating proteins, including for example, Gal4 or

VP16. In one embodiment of the invention, polynucleotides encoding binding partner proteins CREB and CBD are inserted in plasmids pVP16-CREB and pLexA-CBD, respectively, which were deposited with the ATCC and assigned Accession Numbers ATCC 98138 and ATCC 98139, respectively.

The promoter sequence of the invention which regulates transcription of the repressor protein can be any sequence capable of driving transcription in the chosen host cell. The promoter may be a DNA sequence specifically recognized by the chosen DNA binding domain of the invention, or any other DNA sequence with which the DNA binding domain of the fusion protein is capable of high affinity interaction. In a preferred embodiment of the invention, the promoter sequence of the invention is either a HIS3 or alcohol dehydrogenase (ADH) promoter. In a presently most preferred embodiment, the ADH promotor is employed in the invention. The invention, however, encompasses numerous alternative promoters, including, for example, those derived from genes encoding HIS3, ADH, URA3, LEU2 and the like.

In another aspect, the invention provides methods to identify molecules that inhibit interaction between known binding partner proteins. In one embodiment, the invention provides a method to identify an inhibitor of binding between a first binding protein or binding fragment thereof and a second binding protein or binding fragment thereof comprising the steps of (a) growing host cells transformed or transfected as described above in the absence of a test compound and under conditions which permit expression of said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof such that said first binding protein or fragment thereof and second binding protein or binding fragment thereof interact bringing into proximity said DNA binding domain and said transactivating domain forming a functional transcriptional activating protein; the transcriptional activating protein acting on said promoter to increase

expression of said repressor protein; said repressor protein interacting with said operator such that said selectable marker protein is not expressed; (b) confirming lack of expression of said selectable marker protein in said host cell; (c) growing said host cells in the presence of a test compound; and (d) comparing expression of said selectable marker protein in the presence and absence of said test compound wherein increased expression of said selectable marker protein is indicative that the test compound is an inhibitor of binding between said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof.

In a most preferred embodiment, the invention provides a method to identify an inhibitor of binding between a first binding protein or binding fragment thereof and a second binding protein or binding fragment thereof comprising the steps of: (a) transforming or transfecting a host cell with a first DNA expression construct comprising a first selectable marker gene encoding a first selectable marker protein and a repressor gene encoding a repressor protein, said repressor gene under transcriptional control of a promoter; (b) transforming or transfecting said host cell with a second DNA expression construct comprising a second selectable marker gene encoding a second selectable marker protein and a third selectable marker gene encoding a third selectable marker protein, said third selectable marker gene under transcriptional control of an operator, said operator specifically acted upon by said repressor protein such that interaction of said repressor protein with said operator decreases expression of said third selectable marker protein; (c) transforming or transfecting said host cell with a third DNA expression construct comprising a fourth selectable marker gene encoding a fourth selectable marker protein and a first fusion protein gene encoding a first binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activation protein or a transactivating domain of said transcriptional activation protein; (d) transforming or transfecting said host cell with a fourth DNA expression construct comprising

a fifth selectable marker gene encoding a fifth selectable marker protein and a second fusion protein gene encoding a second binding protein or binding fragment thereof in frame with either the DNA binding domain of said transcriptional activation protein or the transactivating domain of said transcriptional activation protein, whichever is not included in first fusion protein gene; (e) growing said host cell under conditions which permit expression of said first binding protein or fragment thereof and said second binding protein or fragment thereof such that said first binding protein or fragment thereof and second binding protein or binding fragment thereof interact bringing into proximity said DNA binding domain and said transactivating domain reconstituting said transcriptional activating protein; said transcriptional activating protein acting on said promoter to increase expression of said repressor protein; said repressor protein interacting with said operator such that said third selectable marker protein is not expressed; (f) detecting absence of expression of said selectable gene; (g) growing said host cell in the presence of a test compound of binding between said first protein or fragment thereof and said second binding protein or fragment thereof; and (h) comparing expression of said selectable marker protein in the presence and absence of said test compound wherein decreased expression of said selectable marker protein is indicative of an ability of the test compound to inhibit binding between said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof such that said transcriptional activating protein is not reconstituted, expression of said repressor protein is not increased, and said operator increases expression of said selectable marker protein.

The methods of the invention encompass any and all of the variations in host cells as described above. In particular, the invention encompasses a method wherein: the host cell is a yeast cell; the selectable marker gene encodes HIS3; transcription of the selectable marker gene is regulated by the *tet* operator; the repressor protein gene encodes the

tetracycline resistance protein; transcription of the tetracycline resistance protein is regulated by the HIS3 promoter; the DNA binding domain is derived from LexA; and the transactivating domain is derived from VP16. In another embodiment, the invention encompasses a method wherein: the host cell is a yeast cell; the selectable marker gene encodes HIS3; transcription of the selectable marker gene is regulated by the *tet* operator; the repressor protein gene encodes the tetracycline resistance protein; transcription of the tetracycline resistance protein is regulated by the alcohol dehydrogenase promoter; the DNA binding domain is derived from LexA; and the transactivating domain is derived from VP16.

In alternative embodiments of the invention wherein the host cell is a mammalian cell, variations include the use of mammalian DNA expression constructs to encode the first and second recombinant fusion genes, the repressor gene, and the selectable marker gene, and use of selectable marker genes encoding antibiotic or drug resistance markers (i.e., neomycin, hygromycin, thymidine kinase).

There are at least three different types of libraries used for the identification of small molecule modulators. These include: (1) chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

Chemical libraries consist of structural analogs of known compounds or compounds that are identified as "hits" via natural product screening. Natural product libraries are collections of microorganisms, animals plants or marine organisms which are used to create mixtures for screening by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of plants or marine organisms. Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds as a mixture. They are relatively easy to prepare by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and

oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, polypeptide libraries.

The utility of the various aspects of the invention is manifest. Host cells of the invention are useful to demonstrate *in vivo* binding capacity of both known and suspected binding partner proteins in a recombinant system. Such an expression system permits systematic analysis of the structure and function of a particular binding protein, thus permitting identification and/or synthesis of potential modulators of the physiological activity of the binding proteins. The methods of the invention are particularly useful to identify and improve molecules which are capable of inhibiting specific and general protein/protein interactions. Inhibitors identified by the methods of the invention can then be examined for utility *in vivo* as therapeutic and/or prophylactic medicaments for conditions associated with various protein/protein interactions.

Description of the Drawing

Figure 1 describes the mechanics of the split hybrid assays.

Detailed Description of the Invention

The present invention relates generally to methods designated split hybrid assays to identify inhibitors of protein/protein interactions and is illustrated by the following examples describing various methods for making and using the invention. In particular, Example 1 relates to construction of various plasmids and expression constructs utilized in the invention. Example 2 described generation of various yeast transformants used to identify inhibitor compounds. Examples 3, 4, 5 and 6 address use of the split hybrid assay to examine CREB/CBD binding, Tax/SRF binding, CKI/CREB binding and AKAP 79 binding to various partner protein, respectively. Example 7 describe general application of the split hybrid assay. Example 8 relates to

use of the split hybrid assay for weakly interacting binding partners. Example 9 describes general assay methods. Example 10 addresses use of the split hybrids assay to identify agents that prevent receptor desensitization and drug tachyphylaxis.

Example 1
Plasmid Construction

In the examples that follow, various plasmid constructs were utilized as described. To simplify discussion of the exemplified assays, this example describes construction of the various plasmids used in the following examples. For clarity, the plasmids are grouped according common features relating to their applications in the assays later discussed.

I. Plasmids Encoding Reporter Gene HIS3

A. pRS303/1xtetop-MluI

One copy of the *tet* operator sequence was engineered into position -53 in the *HIS3* promoter of pRS313 [Sikorski, R.S. *et al.*, *Genetics* **122**:19-27 (1989)] by using the polymerase chain reaction (PCR). Two primary PCR reactions using pRS313 as a template were performed which utilized a 5′-terminal oligonucleotide designated Eco47III-5′ and a 3′-inner oligonucleotide designated Tetop internal 3′ to yield a primary 5′-PCR product and a 5′-inner oligonucleotide designated Tetop internal 5′ and a 3′-terminal oligonucleotide designated Nhe I 3′ to yield a primary 3′-PCR product.

Eco47 III-5′ SEQ ID NO: 1
5′-TTGGTGAGCGCTAGGAGTCACTGCCAG

Tetop int. 3′ SEQ ID NO: 2
5′-TATACTCTATCAATGATAGAGTAATTCATTATGTGATAATGCC

Tetop int. 5′ SEQ ID NO: 3
5′-ATTACTCTATCATTGATAGAGTATATAAAGTAATGTGATTTC)

Nhe I 3′ SEQ ID NO: 4

5'-AATTCTGCTAGCCTCTGCAAAGC

5' and 3' inner oligonucleotides contain complementary sequence such that 3' sequence of the primary 5' PCR product overlaps with 5' sequence of the primary 3' PCR product. The 5' terminal oligonucleotide contains the restriction site *Eco*47III while the 3' terminal oligonucleotide contains the restriction site *Nhe*I in order to facilitate subsequent subcloning. The primary PCR reactions were performed with *Pfu* DNA polymerase (Stratagene, La Jolla, CA) using reaction conditions described by the manufacturer. PCR products were isolated by Bio101 (Vista, CA) Gene Clean III gel extraction. The primary 5' and 3' PCR products were then combined in a second PCR reaction and amplified using the 5'- and 3'- terminal oligonucleotides, Eco47III-5' and Nhe I 3'. The second PCR reaction was performed with *Vent* DNA polymerase (New England Biolabs, Beverly, MA) using reaction conditions described by the manufacturer, except that the reactions were supplemented with 4 mM Mg^{2+}. The final PCR product contained one *tet* operator sequence inserted into position -53 of the *HIS3* promoter and nucleotides 52-48 deleted in the construction. The final PCR product was isolated, digested with *Eco*47III and *Nhe*I and cloned into pRS313 previously digested with *Eco*47III and *Nhe*I. The resulting plasmid was designated pRS313/1xtetop. DNA sequencing confirmed the presence of one copy of the *tet* operator sequence in pRS313/1xtetop and confirmed integrity of the *Eco*47III and *Nhe*I junctions.

A *Mlu*I restriction enzyme site was engineered into position -22 in the *HIS3* promoter of pRS313/1xtetop by utilizing PCR using *Vent* DNA polymerase using pRS313/1xtetop as template. One PCR construct was amplified using the 5' terminal oligonucleotide Eco47 III-5' (SEQ ID NO: 1) containing an *Eco*47III restriction site and a 3'-oligonucleotide designated Mlu I 3' containing a *Mlu*I restriction site.

Mlu I 3′ SEQ ID NO: 5
5′-CGCACGCGTCGAAGAAATCACATTACTTTATATA

A second PCR product was amplified using the 3′-terminal oligonucleotide Nhe I 3′ (SEQ ID NO: 4) containing a *Nhe*I restriction site and a 5′-oligonucleotide designated Mlu I 5′ containing a *Mlu*I restriction site.

Mlu I 5′ SEQ ID NO: 6
5′-CGCACGCGTATACTAAAAAATGAGCAGGCAAG

The first PCR product was isolated and digested with *Eco*47III and *Mlu*I, while the second PCR product was isolated and digested with *Mlu*I and *Nhe*I. These digested products were isolated and ligated in a triple ligation with pRS313 previously digested with *Eco*47III and *Nhe*I. The resulting plasmid was designated pRS313/1xtetop-MluI. DNA sequencing confirmed the presence of the *Mlu*I site in pRS313/1xtetop-MluI and confirmed that integrity of the *Eco*47III and *Nhe*I junctions were maintained.

A pRS303/1xtetop-MluI plasmid was constructed by first removing the *Eco*47III/*Nhe*I fragment containing the altered *HIS3* promoter from the pRS313/1xtetop-*Mlu*I vector and ligating the isolated fragment into pRS303 previously digested with *Eco*47III and *Nhe*I. DNA sequencing confirmed proper insertion of the *Eco*47III/*Nhe*I fragment.

B. <u>pRS303/2xtetop-LYS2</u>

One copy each of the *tet* operator sequence was engineered into positions -53 and -22 in the *HIS3* promoter of pRS303 [Sikorski, *et al.*, *Genetics* 122:19-27 (1989)]. PCR was utilized to engineer one copy into position -53 which resulted in plasmid pRS303/1xtetop. To insert the second copy, a *Mlu*I site was introduced at position -22 in the *HIS3* promoter using PCR. The new plasmid was designated pRS303/1xtetop-MluI.

The *tet* operator was created by annealing two complementary

oligonucleotides tetop-1 and tetop-2.

tetop-1 SEQ ID NO: 7
5'-CGCGTACTCTATCATTGATAGAGTA;

tetop-2 SEQ ID NO: 8
5'-ATGAGATAGTAACTATCTCATGCGC

When annealed, the *tet* operator sequence contains flanking *Mlu*I sites. Both oligonucleotides were phosphorylated using *T4* polynucleotide kinase (Gibco BRL, Grand Island, NY) at 37°C for one hour and annealed by first heating at 70°C for 10 minutes and then cooling to room temperature. The annealed oligonucleotides were isolated and ligated into pRS303/1xtetop-*Mlu*I previously digested with *Mlu*I. The resulting plasmid was designated pRS303/2xtetop. DNA sequencing confirmed insertion of one copy of the *tet* operator sequence in the *Mlu*I site.

The *LYS2* gene was digested from pLYS2 [Hollenberg, S.M. *et al., Mol.Cell.Biol.* **15**:3813-3822 (1995)] with *Eco*RI and *Hind*III and the isolated fragment blunt ended using the large fragment of DNA polymerase I (Gibco BRL, Grand Island, NY). Phosphorylated *Sst*I linkers (New England Biolabs, Beverly, MA) were ligated to the fragment, the fragment digested with *Sst*I, and the resulting fragment ligated into pRS313 previously digested with *Sst*I. The resulting plasmid was designated pRS313/LYS2.

The *LYS2* fragment was removed from pRS313/LYS2 with *Sst*I digestion and inserted into pRS303/2xtetop previously digested with *Sst*I. The resulting plasmid was designated pRS303/2xtetop-LYS2.

Similarly, the *LYS2* *Sst*I fragment was inserted into pRS303/1xtetop-MluI previously digested with *Sst*I yield pRS303/1xtetop-MluI-LYS2.

C. pRS303/3xtetop-LYS2

Two copies of the tet operator sequence were created by self-annealing a palindromic oligonucleotide Tetop 2x with itself.

Tetop 2x SEQ ID NO: 9
5′-CGCGTACTCTATCATTGATAGAGTCTAGACTCTATCAATGATAGAGTA

The annealed oligonucleotide contained flanking *Mlu*I sites. The oligonucleotide was phosphorylated, annealed, and isolated as above. The isolated annealed and *Mlu*I-digested oligonucleotide was ligated into pRS303/1xtetop-*Mlu*I-LYS2 previously digested with *Mlu*I to yield pRS303/3xtetop-LYS2. The presence of two copies of the *tet* operator sequence in the *Mlu*I site was confirmed by DNA sequencing.

D. pRS303/4xtetop-LYS2 and pRS303/8xtetop-LYS2

Three or seven copies of the *tet* operator were created using PCR with *Vent* DNA polymerase as described above. Plasmid pUHC-13-3 [Grossen and Bujarg, *Proc. Natl. Acad. Sci. (USA)* 89:5547-5551 (1992)] was used as template DNA using 5′- and 3′- oligonucleotides, Mlu I/Sph I 5′ and Mlu I Sph I 3′, containing an exterior *Mlu*I restriction enzyme site nested internally by a *Sph*I restriction enzyme site.

Mlu I/Sph 1 5′ SEQ ID NO: 10
5′-GCGACGCGTGCATGCCGTCTTCAAGAATTCCTCGAG

Mlu I Sph I 3′ SEQ ID NO: 11
5′-GCGACGCGTGCATGCCCACCGTACACGCCTACTCGA

The PCR products were separated on an agarose gel and the ladder of different sized DNA fragments was isolated, digested with *Mlu*I, and ligated into the *Mlu*I restriction site of pRS303/1xtetop-MluI-LYS2. DNA sequencing revealed that either three or seven copies of tet operators were inserted into the *Mlu* site of pRS303/1xtetop-*Mlu*I-LYS2 to provide either pRS303/4xtetop-LYS2 or pRS303/8xtetop-LYS2.

E. pRS303/6xtetop-LYS2 and pRS303/10xtetop-LYS2

A *Sph*I restriction enzyme site was introduced at position -85 in the *HIS3* promoter of pRS303/3xtetop-LYS2 using PCR with *Vent* DNA polymerase as described. Plasmid pRS303/3xtetop-LYS2 was used as a template DNA. A first fragment was amplified using the 5'-terminal oligonucleotide Eco47 III-5' (SEQ ID NO: 1) described above containing an *Eco*47III restriction site and a 3'-oligonucleotide Sph I 3' containing a *Sph*I restriction site.

Sph I 3' SEQ ID NO: 12
5'-CATGGCATGCAAAAAAAAAGAGTCATCCGCTAGG

A second PCR product was amplified using the 3'-terminal oligonucleotide Nhe I 3' (SEQ ID NO: 4) described above containing a *Nhe*I restriction site and a 5'-oligonucleotide containing a *Sph*I restriction site.

Sph I 5' SEQ ID NO: 13
5'CATGGCATGCTTAGCGATTGGCATTATCACAT

The PCR products were isolated as described above. The first PCR product was digested with *Eco*47III and *Sph*I, and the second PCR product was digested with *Sph*I and *Nhe*I. Both digestion products were ligated in a triple ligation along with pRS303/3xtetop-LYS2 previously digested with both *Eco*47III and *Nhe*I. The resulting plasmid was designated pRS303/3xtetop-SphI-LYS2. The presence of the *Sph*I site in pRS303/3xtetop-SphI-LYS2 was confirmed by DNA sequencing analysis.

Three copies of *tet* operators were isolated as a single fragment by digesting pRS303/4xtetop-LYS2 with *Sph*I. The isolated fragment was ligated into the *Sph*I site of pRS303/3xtetop-*Sph*I-LYS2 to yield pRS303/6xtetop-LYS2. The presence of three additional copies of the *tet* operator in pRS303/6xtyetop-LYS2 at the *Sph*I site was confirmed by DNA

sequencing.

Seven copies of *tet* operators were isolated as a single fragment by digesting pRS303/8xtetop-LYS2 with *Sph*I. The isolated fragment was ligated into the *Sph*I site of pRS303/3xtetop-*Sph*I-LYS2 to yield pRS303/10xtetop-LYS2. The presence of seven additional copies of the tet operator in pRS303/10xtetop-LYS2 at the *Sph*I site was confirmed by DNA sequencing.

F. pRS313/MluI and pRS303/MluI

A *Mlu*I restriction enzyme site was engineered into position -22 in the *HIS3* promoter of pRS313 utilizing PCR and *Vent* DNA polymerase as noted above. Plasmid pRS313 was used as a template for these PCR reactions. One PCR construct was amplified using the 5´ terminal oligonucleotide Eco47 III-5′ (SEQ ID NO: 1) containing an *Eco47*III restriction site and a 3´ oligonucleotide Mlu I 3′ (SEQ ID NO: 5) containing a *Mlu*I restriction site. A second PCR product was amplified using the 3´ terminal oligonucleotide Nhe I 3′ (SEQ ID NO: 4) containing a *Nhe*I restriction site and the 5´ oligonucleotide Mlu I 5′ (SEQ ID NO: 6) containing a *Mlu*I restriction site. The first PCR product was isolated and digested with *Eco47*III and *Mlu*I, while the second PCR product was isolated and digested with *Mlu*I and *Nhe*I. The digested products were partially purified and joined in a triple ligation with pRS313 which had been previously digested with *Eco47*III and *Nhe*I. The resulting plasmid was designated pRS313/MluI. DNA sequencing confirmed the presence of the *Mlu*I site in pRS313/MluI and to confirm the integrity of the *Eco47*III and *Nhe*I junctions.

pRS303/MluI was constructed in exactly the same manner as pRS313/MluI except that pRS303 was used in place of pRS313.

G. pRS313/1xtetop

See above wherein pRS313/1xtetop is an intermediate in the

construction of pRS303/1xtetop-MluI.

H. pRS313/MluI-1xtetop and pRS303/MluI-1xtetop

One copy of the *tet* operator sequence was created by annealing two complementary oligonucleotides tetop-1 and tetop-2 (SEQ ID NO: 7 and SEQ ID NO: 8). The annealed *tet* operator sequence contains flanking *Mlu*I sites. The oligonucleotides were phosphorylated using *T4* polynucleotide kinase (Gibco BRL, Grand Island, NY) at 37°C for one hour and annealed by first heating at 70°C for 10 minutes followed by cooling to room temperature. The annealed oligonucleotides were isolated and ligated separately into *Mlu*I-digested pRS313/MluI and pRS303/MluI, the resulting plasmids being designated pRS313/MluI-1xtetop and pRS303/MluI-1xtetop. DNA sequencing confirmed the presence of one copy of the *tet* operator in the *Mlu*I sites of both plasmids.

In order to produce plasmids bearing multiple copies of the *tet* operator, annealed oligonucleotides described above were ligated together overnight at 16°C. After isolation of the ligation products, they were inserted into the *Mlu*I of pRS313/MluI. DNA sequencing analysis confirmed that one clone, pRS313/MluI-4xtetop, was produced which contained four copies of *tet* operator in the *Mlu*I site. However, upon further examination of this clone it was discovered that it had been subjected to a recombination event and was therefore not useful for further cloning steps. Continued attempts to insert multiple copies of the *tet* operator into the *Mlu*I site of pRS313/MluI by ligating multimers of the *tet* operator have been unsuccessful.

I. pRS313/1xtetop-MluI

See above wherein construction of pRS313/1xtetop-MluI was an intermediate in the construction of pRS303/1xtetop-MluI.

J. pRS313/2xtetop

One copy of the *tet* operator sequence was created using annealed complementary oligonucleotides tetop-1 and tetop-2 (SEQ ID NO: 7 and SEQ ID NO: 8). Annealed oligonucleotides were ligated into the *MluI* site of pRS313/1xtetop-MluI to yield pRS313/2xtetop. DNA sequencing confirmed the presence of two copies of the *tet* operator in the *Mlu*I site.

K. pRS303/2xtetop

See above wherein pRS303/2xtetop was an intermediate in the construction of pRS303/2x/tetop-LYS2.

L. pRS313/LYS2 and pRS313/LYS2

The *LYS2* gene was digested from pLYS2 with *Eco*RI and *Hin*dIII digestion. The *Eco*RI/*Hin*dIII fragment was blunt ended using the large fragment of DNA polymerase I (Gibco BRL, Grand Island, NY) and ligated with phosphorylated *Sst*I linkers (New England Biolabs, Beverly, MA). The resulting fragment was digested with *Sst*I and ligated into pRS313 previously digested with *Sst*I. The resulting plasmid was designated pRS313/LYS2. Because the *LYS2* fragment was shown to have inserted into pRS313 in both orientations, plasmids with the *LYS2* gene in both orientations were transformed separately into the yeast strain SEY6210α_(*MAT*α_ *leu2-3,112 ura3-52 his3-*Δ*200 trp1-*Δ*901 lys2-801 suc2-*Δ*9* [Robinson *et al.*, *Mol. Cell. Biol.* 8:4936-4948 (1988)]. Both clones allowed the yeast to grow in the absence of lysine indicating that orientation of the *LYS2* gene in pRS313 did not affect the expression of an active gene.

The *LYS2* fragment was removed from pRS313/LYS2 with *Sst*I and ligated into the *Sst*I site of:

pRS313/1xtetop-MluI giving plasmid pRS313/1xtetop-MluI-LYS2,
pRS313/2xtetop giving plasmid pRS313/2xtetop-LYS2,

pRS303/1xtetop-MluI giving plasmid pRS303/1xtetop-MluI-LYS2, and pRS303/2xtetop giving plasmid pRS303/2xtetop-LYS2.

II. Plasmids Encoding Reporter Gene TetR

A. pRS306/HIS3:TetR/Term

The 5′ promoter sequence of the yeast *HIS3* gene, encompassing nucleotides -75 to +23, was ligated to the translational start of *TetR*. In addition, the DNA sequence encoding the simian virus 40 (SV40) large T antigen nuclear localization signal was ligated in frame with the nucleotide sequence encoding the last amino acid residue of TetR. The chimeric fragment was created by the same PCR strategy as described above.

The *HIS3* promoter fragment, the primary 5′-PCR product, was amplified by PCR from plasmid p601 [Grueneberg,D.A., *Science* 257:1089-1095 (1992)] using a 5′-terminal oligonucleotide T7 Promoter primer and a 3′-inner oligonucleotide 3′-TetR inner primer.

T7 Promoter primer SEQ ID NO: 14
5′-TAATACGACTCACTATATAGGG

3′-TetR inner primer SEQ ID NO: 15
5′-TCTAGACTTTGCCTTCGTTTATC

The primary 3′ PCR product containing the *TetR* coding sequence was amplified from pSLF104 [Forsburg, *Nucl. Acid. Res.* 21:2955-2956 (1993)] with a 5′-inner oligonucleotide 5′-TetR inner primer and a 3′-terminal oligonucleotide 3′-TetR terminal primer.

5′-TetR inner primer SEQ ID NO: 16
5′-CGAAGGCAAAGATGTCTAGATTAGATAAAAG

3′-TetR terminal primer SEQ ID NO: 17
5′-CGCGGATCCGCTTTCTCTTCTTTTTTGGAGACCCACTTTCACATTTAAG

An *Eco*RI site derived from the p601 fragment and a *Bam*HI site in the 3'-terminal oligonucleotide were used in subsequent subcloning. The PCR products were gel-purified and amplified in a second PCR reaction with 5'- and 3-' terminal oligonucleotides, T7 Promoter primer (SEQ ID NO: 14) and 3'-TetR terminal primer (SEQ ID NO: 17). The secondary PCR product was isolated, digested with *Eco*RI and *Bam*HI, and ligated into pRS306/Term previously digested with *Eco*RI and *Bam*HI. The resulting plasmid was designated pRS306/HIS3:TetR/Term which comprises the complete TetR coding sequence in frame with sequences encoding the nuclear localization signal of SV40 large T antigen.

B. pRS316/HIS3:TetR/Term

The construction protocol for this plasmid was the same as described above for subcloning a *HIS3* DNA into pRS306/Term except that the vector for subcloning was pRS316/Term described above.

C. pRS306/1xLexAop/HIS3:TetR

Oligonucleotides LexAop (100a) and LexAop (100b) containing a single copy of LexA operator were phosphorylated with *T4* polynucleotide kinase (Gibco BRL, Grand Island, NY) at 37°C for one hour.

LexAop (100a) SEQ ID NO: 18
5'-AATTGCTCGAGTACTGTATGTACATACAGTAG

LexAop (100b) SEQ ID NO: 19
5'-AATTCTACTGTATGTACATACAGTACTCGAGC

Following phosphorylation, the oligonucleotides were annealed by heating at 70°C for 10 minutes followed by cooling to room temperature. The annealed oligonucleotide containing 5´ and 3´ *Eco*RI overhanging ends was subcloned into pRS306/HIS3:TetR/Term previously digested with *Eco*RI. The number

of copies of inserted oligonucleotide was confirmed by DNA sequencing. The plasmid containing a single copy of the LexA operator was designated pRS306/1xLexAop/HIS3:TetR.

D. pRS316/2xLexAop/HIS3:TetR

The subcloning protocol for this construct was the same as described above for pRS306/1xLexAop/HIS3:TetR. The annealed oligonucleotides encoding the LexA operator included overhanging *Eco*RI ends and during ligation, the individual annealed fragments were able to multimerize, inserting into the parental plasmid more than one copy of the desired LexA sequence. The number of copies of inserted oligonucleotides was confirmed by DNA sequencing.

E. pRS306/2xLexAop/HIS3:TetR

A DNA fragment containing two copies of LexA operator and the chimeric *HIS3*:TetR reporter was excised from pRS316/2xLexAop/HIS3:TetR by digestion with *Kpn*I and *Bam*HI restriction enzymes. The fragment was gel-purified and subcloned into pRS306/Term previously digested with *Kpn*I and *Bam*HI and the resulting construct was sequenced to confirm the presence of two copies of the LexA operator.

F. pRS306/4xLexAop/HIS3:TetR
and pRS306/8xLexAop/HIS3:TetR

A pair of oligonucleotides SH101A and SH101B were utilized in PCR to amplify the LexA binding site multimer from the plasmid SH18-34ΔSpe [Hollenberg, S.M., *et al.*, *Mol.Cell.Biol.* **15**:3813-3822 (1995)].

SH101A SEQ ID NO: 20
5'-CCGGAATTCTCGAGACATATCCATATCTAATC

SH101B SEQ ID NO: 21
5'-CCGGAATTCACTAATCGCATTATCATC

The amplification product containing four copies of LexA operator was gel-purified, digested with *Eco*RI, and subcloned into pRS306/HIS3:TetR/Term previously digested with *Eco*RI. The number of LexA operators were determined by DNA sequencing.

G. pRS306/8xLexAop/HIS3::TetR

A PCR strategy was used to link the 5′ promoter sequence of the yeast *HIS3* gene encompassing nucleotides-75 to +23 to the translational start of *TetR*. Sequences encoding the SV40 large T antigen nuclear localization signal were fused in frame with the nucleotide sequence encoding the last amino acid residue of TetR. The PCR product was digested with *Eco*RI and *Bam*HI and inserted into pRS306/Term previously digested with *Eco*RI and *Bam*HI. The resulting plasmid was designated pRS306/HIS3:TetR/Term, and was shown to encode the complete TetR protein in frame with the nuclear localization signal of SV40 large T antigen. The fusion protein is followed by four amino acids generated by the vector backbone (Arg-Ile-His-Asp).

The LexA binding site multimer from the plasmid pSH18-34ΔSpe [Hollenberg, S.M. *et al.*, *Mol.Cell.Biol.* **15**:3813-3822 (1995)] was amplified by PCR, digested with *Eco*RI, and subcloned into the *Eco*RI site of pRS306/HIS3:TetR/Term resulting in plasmid pRS306/8xLexAop/TetR.

H. pADH/TetR

The DNA coding sequence of TetR was amplified by PCR from pSLF104 using two oligonucleotides, NcoI-TetR and 3′-TetR terminal primer (SEQ ID NO: 17).

NcoI-TetR SEQ ID NO: 22

5′-CATGCCATGGCCATGTCTAGATTAGATAAAAG

The resulting product was gel-purified, digested with *Nco*I and *Bam*HI, and subcloned into a pBTM116 [Bartel, *et al.*, in Cellular Interactions in Development: a Practical Approach, Hartley (ed.), IRL Press; Oxford, pp. 153-179 (1993)] shuttle vector containing an ADH promoter, previously digested with *Nco*I and *Bam*HI. For construction of this vector, DNA generated by PCR and DNA obtained by restriction enzyme digestion of the polylinker region in plasmid pBluescript (Stratagene, La Jolla, California) were used to engineer additional restriction sites 5′ and 3′ of the ADH promoter. The TetR protein encoded from this construct is expressed containing additional amino acids Met^{-2}-Ala^{-1} before the initiating methionine and also contains the nuclear localization signal of SV40 large T antigen located after the last amino acid of TetR as described above.

I. pRS306/ADH:TetR/Term

A fragment encoding the ADH promoter and TetR was removed from plasmid pADH/TetR with *Xho*I and blunted-ended with the large fragment of DNA polymerase I (Gibco BLR, Grand Island, NY). *Eco*RI linkers (New England BioLabs, Beverly, MA) were added and the fragment was digested with *Eco*RI and *Bam*HI. The resulting fragment was gel-purified and ligated into pRS306/Term previously digested with *Eco*RI and *Bam*HI.

J. pRS306/4xLexAop/ADH::TetR and pRS306/8xLexAop/ADH::TetR

The subcloning protocol used to insert multiple copies of the LexA operator into pRS306/ADH:TetR/Term was the same as described previously for pRS306/4xLexAop/HIS3:TetR and pRS306/8xLexAop/HIS3:TetR.

III. Plasmids Encoding Binding Proteins

A. pLexA-CBD

A DNA fragment containing the CREB binding domain of CBP (CBD), amino acids 461-682, was PCR amplified from plasmid CBP-0.8 [Chrivia, J.C. *et al.*, *Nature* 365:855-859 (1993)] using a pair of oligonucleotides designated 5′ CBD primer and 3′ CBD primer.

5′ CBD primer SEQ ID NO: 23
5′-GCGAATTCGCCAGGGCAACAGAATGCCACT

3′ CBD primer SEQ ID NO: 24
5′-CGGGATCCTGGCTGGTTACCCAGGATGCCTTG

Following gel purification, the amplification product was digested with *Eco*RI and *Bam*HI, and ligated into plasmid pBTM116 [Bartel, *et al.*, in *Cellular Interactions in Development: a Practical Approach*, (ed) Hartley, D.A. (IRL Press, Oxford), pp. 153-179 (1993)] previously digested with *Eco*RI and *Bam*HI.

B. pVP16-CBD

A DNA fragment encoding the CBP sequence was excised from pLexA-CBD by digestion with *Eco*RI and *Bam*HI. Plasmid pLexA-CBD was linearized with *Eco*RI digestion, the resulting overhanging ends blunt-ended using the Klenow fragment of DNA polymerase I, and the ends ligated with *Bam*HI linkers. The resulting fragment was inserted into pVP16 [Hollenberg, *et al.*, *Mol. Cell. Biol.* 15:3813-3822 (1995)] previously digested with into *Bam*HI.

C. pVP16 CREB

Plasmid pcDNA3/CREB283 [Sun and Maurer, *J. Biol. Chem.* 270:7041-7044 (1995)], containing the VP16 transactivation domain fused to sequences of the rat CREB transactivation domain (1 to 283 aa) was linearized with *Xho*I and *Bam*HI linkers (New England BioLab) ligated to the resulting blunt-ended *Xho*I sites. DNA encoding the VP16/CREB chimeric protein was removed with *Hind*III and *Bam*HI digestion and following gel purification, ligated into the *Hind*III and *Bam*HI sites of pVP16 which encodes the *LEU2* gene.

D. pVP16-CREB(BglII-SacII)-LacZ

A DNA fragment encoding β-galactosidase was PCR amplified from plasmid pSV-β-galactosidase vector (Promega, Madison, WI) using a pair of oligonucleotides, 5′ β-gal primer and 3′ β-gal primer and inserted into the *Not*I site of pVP16 to produce pVP16-LacZ.

5′ β-gal primer SEQ ID NO: 29
5′-ATGGTACCAGCGGCCGCTAGTCGTTTTACAACGTCGTGAC

3′ β-gal primer SEQ ID NO: 30
5′-ATGGTACCGCGGCCGCTTATTTTTGACACCAGACCAAC

A PCR fragment containing CREB sequences encoding amino acid residues 1 to 283 was amplified from plasmid pRSV-CREB341 [Kwok, *et al.*, *Nature* 380: 642-646 (1996)] using a pair of oligonucleotides, 5′ CREB 341 primer and 3′ CREB 283 primer, and inserted into pVP16-LacZ vector at the *Bam*HI site.

5′ CREB 341 primer SEQ ID NO: 25
5′-CGCGGATCCGGATGACCATGGACTCTGGAG

3′ CREB 283 primer SEQ ID NO: 28
5′-CGCGGATCCGTGCTGCTTCTTCAGCAGGCTG

To generate a cassette vector for producing and subcloning mutated CREB sequences as described below, PCR was used to engineer a *Bgl*II site using oligonucleotides 5´ *Bgl*II primer and 3´ *Bgl*II primer, at nucleotides 273 to 278 and a *Sac*II site using oligonucleotides 5' *Sac*II primer and 3' *Sac*II primer at nucleotides 500 to 505 of the CREB activation domain.

5´ *Bgl*II primer SEQ ID NO: 31
5´-CGGAGATCTAAAGAGACTTTTCTCCGGAACTCAG

3´ *Bgl*II primer SEQ ID NO: 32
5´-CGGAGATCTTTACAGGAAGACTGAACTGT

5 *Sac*II primer SEQ ID NO: 33
5´-CCACCGCGGCAGTGCCAACCCCGATTTAC

3´ *Sac*II primer SEQ ID NO: 34
3´-CATCCGCGGTGGTGATGGCAGGGGCTGA

E. pLexA-CREB 283

A DNA fragment containing the rat CREB transactivation domain (amino acids 1 to 283) was excised from pcDNA/CREB283 [Sun and Maurer, *supra*] with *Sma*I and *Xba*I digestion. The 5´ *Xba*I site was blunt ended with the large fragment of DNA polymerase I (Gibco BRL, Grand Island, NY) and *Sal*I linkers (New England Biolabs, Beverly, MA) added. The fragment was digested with *Sal*I and subcloned into the *Sal*I site of pBTM116.

F. pLexA-CREB 341

A DNA fragment containing the rat CREB 341 cDNA was amplified by PCR from pcDNA/CREB341 [Kwok, *supra*] using a pair of oligonucleotides, 5´ CREB 341 primer (SEQ ID NO: 25) and 3´ CREB 341 primer.

3´ CREB 341 primer SEQ ID NO: 26
5´-CGCGGATCCTTAATCTGACTTGTGGCAGTA

After gel purification, the PCR product was digested with *Bam*HI, and subcloned into the *Bam*HI site of pBTM116.

G. pLexA-CREB 341-M1

A DNA fragment containing the rat CREB sequence with a mutation changing serine at position 133 to alanine was amplified by PCR from plasmid Rc/RSV CREB-M1 [Kwok, *et al., supra*] using the same set of primers as described for pLexA-CREB 341, 5´ CREB 341 primer (SEQ ID NO: 25) and 3´ CREB 341 primer (SEQ ID NO: 26). The resulting amplification product was gel-purified, digested with *Bam*HI, and subcloned into the *Bam*HI site of pBTM116.

H. pVP16-CREB M1

A PCR fragment containing CREB sequences coding for amino acid residues 1 to 283 including the serine 133 mutation to alanine was amplified using a pair of oligonucleotides, 5´ CREB 283 primer and 3´ CREB 283 primer (SEQ ID NO: 28). The PCR fragment was gel-purified, digested with *Bam*HI and inserted into the *Bam*HI site of pVP16.

5´ CREB 283 primer SEQ ID NO: 27
5´-CGCGGATCCCCATGACCATGGAATCTGGAGCC

I. pLexA-SRF

A DNA fragment containing human SRF was excised from plasmid pCGN-SRF [Grueneberg, D.A., *et al.*, *Science*, 257:1089-1095 (1992)] with *Xho*I and *Bam*HI digestion. The *Xho*I site of the fragment was blunt-ended by the large fragment of DNA polymerase I (Gibco BRL, Grand Island, NY), ligated with *Bam*HI linkers, digested with *Bam*HI, and inserted

into pBTM116 previously digested with *Bam*HI.

J. pVP16-Tax

A DNA sequence encoding full length Tax protein was excised from pS6424 [Kwok, R.P.S., *et al.*, *Nature* 380:642-646 (1996)] with *Bam*HI digestion and was inserted into pVP16 previously digested with *Bam*HI.

IV. Plasmids For Binding Protein Controls

A. pLeu

Plasmid pVP16 was digested with *Hin*dIII and *Bam*HI to remove the fragment encoding the VP16 transactivation domain. The digested vector was blunt-ended and self-ligated.

B. pLexA-VP16

The VP16 transactivation domain was PCR amplified from pGal-VP16 [Sadowski, *et al.*, *Nature* 335:563-564 (1988)] with a pair of oligonucleotides, 5´-VP16SH and 3´VP16SH and the resulting amplification product was digested with *Cla*I, blunt-ended, and inserted into pBTM116.

5´-VP16SH SEQ ID NO: 35
GGCTATCGATACGGCCCCCCCGACCGAT

3´-VP16SH SEQ ID NO: 36
GCGTATCGATCTACCCACCGTACTCGTC

C. pLexA-Lamin

See Hollenberg, S.M. *et al.*, *Mol.Cell.Biol.* 15:3813-3822 (1995)].

V. Plasmids Encoding Reporter Gene Controls

A. pRS306/Term

The alcohol dehydrogenase (ADH) terminator sequence was excised from plasmid pBTM116 [Bartel, *et al.*, in *Cellular Interactions in Development: a Practical Approach*, (ed) Hartley, D.A. (IRL Press, Oxford), pp. 153-179 (1993)] with *Sph*I and *Pst*I restriction enzymes and both 3'-overhanging sequences were blunted by *T4* DNA polymerase (Gibco BLR, Grand Island, NY). The fragment was gel-purified and subcloned into the blunt-ended *Not*I site in pRS306 [Sikorski and Hieter, *Genetics*:122:19-27 (1989)]. The orientation of inserted fragment was determined by DNA sequencing.

B. pRS316/Term

The subcloning protocol for inserting the ADH terminator sequence into pRS316 was the same as described for inserting the ADH sequence in pRS306.

Example 2
Generation of Yeast Assay Transformant

Selection of an appropriate yeast assay strain is an empirical determination based on growth characteristics of the transformed alternatives. A general method to make the appropriate selection is described as follows.

Candidate yeast assay strains were transformed individually with reporter gene constructs and/or a plasmid encoding one of the experimental binding proteins. Assay strains thus transformed were then compared for relative differences in growth characteristics, with an optimal assay strain showing negligible growth on media lacking histidine and vigorous growth on media containing histidine. In practical application of this first step in selection using various plasmids transformed into assay strain YI584, the following results were observed.

When the plasmid pLexA-VP16 encoding both the LexA DNA binding domain and the VP16 transactivating domain as a single protein was introduced into the assay cells, growth in the absence of histidine in the media was significantly reduced three days after transformation.

In assays including transformation with plasmids encoding multiple copies of the *tet* operator upstream of the *HIS3* gene, the following plasmids were separately utilized:

pRS303/1xtetop-*HIS* (encoding a single *tet* operator sequence),
pRS303/2xtetop-*HIS* (encoding two *tet* operator sequences),
pRS303/3xtetop-*HIS* (encoding three *tet* operator sequences),
pRS303/4xtetop-*HIS* (encoding four *tet* operator sequences),
pRS303/6xtetop-*HIS* (encoding six *tet* operator sequences),
pRS303/8xtetop-*HIS* (encoding eight *tet* operator sequences), or
pRS303/10xtetop-*HIS* (encoding ten *tet* operator sequences).

In the assay strains transformed with plasmids encoding either one, two, or three copies of the *tet* operator upstream from the *HIS3* gene, cells grew on media lacking histidine at a rate similar to cells grown on media containing histidine. In yeast assay strains transformed with plasmids encoding either six, eight, or ten copies of the *tet* operator upstream from the *HIS3* gene, cell growth was low suggesting that these strains would not be useful in assays to examine binding and interruption of binding between test proteins. These results suggested that, in assay strains transformed with a reporter plasmid having more than three *tet* operator sequences upstream from the *HIS3* gene, normal activity of the *HIS3* promoter is disrupted and that these plasmids would not be useful.

In assays wherein yeast cells were transformed with only reporter plasmids (and not plasmids encoding binding partner fusion proteins) encoding multiple copies of the LexA operator 5′ of the TetR gene, the

following results were observed. Growth of assay cells transformed with plasmids bearing one, two, four, and eight copies of the regulatory LexA operator upstream of the TetR gene appeared to be "copy number" dependent. Yeast cells transformed with plasmids having two copies of the LexA operator grew at a rate significantly higher than those assay cell transformed with a plasmid bearing only one copy of the operator. Cells transformed with plasmids encoding either four or eight LexA operators upstream of the TetR gene grew at an approximately equal rate, and better than assay cells bearing a TetR gene driven by two copies of the operator.

When the alcohol dehydrogenase (ADH) promoter was included upstream of the LexA operator (plasmids encoding either four or eight LexA operators) in the various reporter gene constructs, cell viability was the lowest.

The various cell lines constructed by the methods described above are shown in Table 1, wherein various transformed yeast strains are identified (Strain #) along with the number of LexA operator sequences in the plasmid encoding TetR, the number of tetracycline operator sequences regulating expression of HIS3, and relative growth rate of the transformed strain on media containing histidine. It is important to note that growth variation of transformed cells in media containing histidine is observed, even in cell lines identically transformed. The number of "+" signs in Table 1 is indicative of the host cell's relative ability to grow on media lacking histidine in the absence of transformation with plasmids encoding potential binding proteins. Also in Table 1, a subscript "a" is indicative of transformation with a plasmid bearing the alcohol dehydrogenase promoter; absence of a subscript "a" indicates use of the HIS3 promoter.

Table 1

Various Yeast Transformants

Diploids L40 Strain #	LexA	TetOp	His+
Y1579	1X	2X	+++
Y1581	1X	2X	+++
Y1580	2X	2X	+++
Y1582	2X	2X	+++
Y1583	4X	2X	+++
Y1585	4X	2X	+++
Y1587	4X	2X	+++
Y1589	4X	2X	+++
Y1584	8X	2X	+++
Y1586	8X	2X	+++
Y1588	8X	2X	+++
Y1590	8X	2X	+++
Y1591	2X	2X	+++
Y1594	2X	2X	+++
Y1597	2X	4X	--
Y1633	2X	4X	+
Y1636	2X	4X	++
Y1600	2X	6X	--
Y1606	2X	6X	+
Y1630	2X	6X	+
Y1627	2X	6X	+++
Y1603	2X	10X	+
Y1621	2X	10X	++
Y1609	2X	10X	+
Y1624	2X	10X	++
Y1593	$4X_a$	2X	+++
Y1595	4X	2X	+++
Y1599	$4X_a$	4X	--
Y1634	4X	4X	+
Y1638	$4X_a$	4X	+

Diploids L40 Strain #	LexA	TetOp	His+
Y1602	$4X_a$	6X	--
Y1607	4X	6X	+++
Y1628	4X	6X	+++
Y1632	$4X_a$	6X	- ?
Y1605	$4X_a$	10X	--
Y1610	4X	10X	+
Y1622	4X	10X	++
Y1626	$4X_a$	10X	++
Y1592	8X	2X	+++
Y1596	$8X_a$	2X	+++
Y1598	8X	4X	+
Y1635	$8X_a$	4X	+
Y1637	8X	4X	++
Y1601	8X	6X	+
Y1608	$8X_a$	6X	+
Y1629	$8X_a$	6X	+++
Y1631	8X	6X	+++
Y1604	8X	10X	+
Y1611	$8X_a$	10X	+
Y1623	$8X_a$	10X	++
Y1625	8X	10X	++

Strain#	LexA	TetOp	strain	His+
Y1664	$4X_a$	3X	w303(50)	+++
Y1666	$4X_a$	3X	w303(51)	+++
Y1668	$4X_a$	2X	L40 (69)	+++
Y1670	$4X_a$	2X	L40 (70)	+++
Y1665	$8X_a$	3X	w303(50)	+++
Y1667	$8X_a$	3X	w303(51)	+++
Y1671	$8X_a$	3X	L40 (69)	+++
Y1669	$8X_a$	2X	L40 (69)	+++
Y1671	$8X_a$	2X	L40 (70)	+++
Y1671	$8X_a$	6X	L40 (69)	+++

Example 3
CREB/CBP Binding Interaction

Use of the split-hybrid assay for studies of protein/protein binding wherein one of the binding components is randomly mutagenized was carried out using CREB and CBP binding proteins. The binding of CREB to CBP has been shown to require the phosphorylation of the CREB serine residue at position 133 in a region designated the "kinase-inducible domain" (KID) [Chrivia, *et al.*, *Nature* **365**, 855-859 (1993); Kwok, *et al.*, *Nature* **370**, 223-226 (1994)]. Functionally, changing serine at position 133 to alanine (a mutant designated CREB-M1) abolishes the ability of CBP to activate CREB-mediated transcription. Preliminary studies have indicated that the CREB-M1 mutant in the split-hybrid system prevents the interaction with CBP and subsequent growth of the yeast assay strain on media lacking histidine. Precisely what other portions of the KID of CREB are required for binding to CBP is unknown, however. To define other potentially important amino acid residues, the KID (amino acid residues 102 to 160) of CREB 341 was randomly mutagenized using PCR.

A. PCR Mutagenesis and Creation of Mutant Library

The technique used for mutagenic PCR was a modification of that described by Uppaluri and Towle [*Mol. Cell. Biol.* **15**, 1499-1512 (1995)]. The reaction mixture contained 20 ng of pVP16-CREB(BglII-SacII)-LacZ, 16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl, pH 8.8, 6.1 mM $MgCl_2$, 0.5 mM $MnCl_2$, 6.7 μM EDTA, 10 mM β-mercaptoethanol, 1 mM primers, 1mM each dGTP, dTTP, and dCTP, 400 μM dATP, and 2.5 units of *Taq* DNA polymerase (Promega, Madison, WI). After seven cycles of PCR (94°C for 40 sec, 50°C for 40 sec, and 72°C for 40 sec), the PCR product was amplified a second time using the same primers and *Vent* DNA polymerase (New England BioLabs, Beverly, MA) under the same conditions for 25 cycles. The resultant PCR product was gel purified, digested with *Bgl*II and

*Sac*II, and inserted into the *Bgl*II and *Sac*II sites of pVP16-CREB(BglII-SacII)-LacZ (construction of which is described above). The resulting plasmids were transformed into DH5α bacterial cells. Transformants were pooled and plasmid DNA was isolated by CsCl gradient centrifugation.

B. Construction and Use of pVP16-CREB(BglII-SacII)-LacZ

A DNA fragment encoding the β-galactosidase gene was fused in frame to the carboxyl-terminal end of VP16-CREB as described above. The carboxy-terminal tag allowed identification of clones that contain frame-shift and nonsense mutations; colonies that remain positive for β-galactosidase were presumed to contain an open reading frame throughout the mutated region. To facilitate the subcloning of mutated sequences, a cassette version of the CREB cDNA was generated that contained *Bgl*II and a *Sac*II sites flanking the 5´ and 3´ ends of the KID, respectively. These modifications altered the amino acid residue at position 168 from valine to alanine. The cDNA altered in this manner was indistinguishable from the original VP16-CREB and from VP16-CREB-LacZ when tested in the split hybrid assay. Primers complementary to regions flanking the KID were used in mutagenic PCR amplification reactions as described above under conditions which were optimized to achieve one to three mutations in the 177 bp region encoding the KID. PCR products were introduced into pVP16-CREB(*Bgl*II-*Sac*II)-LacZ in place of wild-type sequence. A library of mutated sequences was transformed into yeast assay strain YI584 expressing LexA-CBD. Approximately 27,000 yeast transformants were screened, yielding about 5,000 colonies that were capable of growing on selective media supplemented with 10 μg/ml of tetracycline and 1 mM of 3AT, determined as described below.

Two screening steps were performed to eliminate uninformative mutations and false positives. First, filter β-galactosidase assays were performed by standard methods [Vojtek, *et al.*, *Cell* 74:205-214 (1993)] on the 5,000 colonies which exhibited positive growth on media lacking

tryptophan, histidine, uracil, leucine, and lysine to eliminate expressed proteins having frame-shift and nonsense mutations. Five hundred thirty six colonies developed a dark blue color, whereas 412 colonies turned white and were presumed to express mutants containing either frame-shift or nonsense mutations. The other colonies developed a pale blue color, and control experiments suggested that these colonies may have expressed unstable lacZ fusion proteins. Pale blue colonies were not analyzed further.

DNA from 536 dark blue colonies was isolated and transformed into *E.coli* MC1066 cells. One hundred ninety three pVP16-CREB-(BglII-SacII)-LacZ cDNAs were then isolated.

In a second screening step, the 193 cDNAs were separately re-transformed along with pLexA-CBD into the split-hybrid strain as well as into the two-hybrid L40 strain [Vojtek, *et al.*, *supra*] in order to identify false positives and confirm that the mutant CREB proteins did not interact with CBP. Among the 193 cDNAs re-screened, 152 did not interact with CBP in the yeast two-hybrid system, 15 interacted weakly, and 26 interacted like wild type CREB.

Following these two screening steps, the 152 CREB mutants were sequenced. Seventy CREB mutants were found to contain a single amino acid change. Sixty four CREB mutants contained two amino acid residue mutations and 13 mutants contained more than two amino acid mutations. Mutants containing more than one amino acid alteration were not analyzed further. The expression level of mutant proteins having one amino acid change were determined using a standard β-galactosidase assay.

The CREB mutations identified in the split-hybrid screen were shown to carry amino acid changes centered around the phosphorylation site at serine at position 133. No disrupting mutations were found to contain amino acid alterations outside of the region between amino acids 130 to 141. Most of the mutations abrogated the PKA phosphorylation region, but others were identified at isoleucine position 137, leucine at position 138, and leucine

at position 141. The mutations at positions 137, 138, and 141 generally changed the hydrophobic residues at these positions to polar residues. The ability of the split-hybrid system to detect only a limited number of CREB mutants, many of which have been proposed previously to disrupt CREB association with CBP [Parker, *et al.*, *Mol. Cell. Biol.* **16**, 694-703. (1996)], indicates the specificity of the split-hybrid system.

These results lead to interesting suggestions. Various CREB mutations were identified which disrupt CREB-CBP interaction and the majority of disrupting mutations occurred in the CREB PKA phosphorylation motif. This result was consistent with previous observations that nonphosphorylated CREB and CBP do not interact [Kwok, *et al.*, *Nature* 370:223-226 (1994)]. The most common motif for PKA phosphorylation is an RRX(S/T)X amino acid sequence but RX(S/T)X and KRXX(S/T)X are also phosphorylated [Kemp and Pearson, *T.I.B.S.* **15**, 342-346 (1990)]. The arginine residues in the phosphorylation site are critical for electrostatic interactions with acidic amino acid residues in the catalytic subunit of PKA [Knighton, *et al.*, *Science* **253**, 414-420 (1991)], and consistent with this observation, CREB mutants with changes at arginine residues 130 and 131 were identified in the split hybrid assay that did not interact with CBP.

Results also showed that CREB mutations at amino acids proline at residue 132 and tyrosine 134 were unable to bind CBP. It is likely that the mutations at these residues adversely affect the structure of the phosphorylation motif, although these positions are generally thought to be less critical to CBP binding. It is possible that the substitution of proline at position 132 with threonine created a new phosphorylation site (RXTX) that interfered with the critical phosphorylation of serine at position 133. Although not generally thought to be part of the "classical" consensus PKA phosphorylation motif, hydrophobic amino acids are commonly found carboxy-terminal to PKA sites [Kemp, *et al.*, *T.I.B.S.* **19**:440-444 (1994)]. The importance of these flanking residues may explain the frequent occurrence

of disrupting mutations involving tyrosine at position 134. Further studies will be directed to determining if mutations of proline at position 134 and tyrosine at position 134 directly disrupt phosphorylation of serine at position 133 or disrupt binding of CREB to CBP by some other mechanism.

In addition, substitution of serine at position 133 with threonine also prevented the interaction of CREB and CBP. PKA protein substrates containing a phosphorylatable threonine residue are known to exist in nature (*i.e.*, protein phosphatase inhibitor 1 and myelin basic protein), although they are less common than those with phosphorylatable serines [Zetterqvist, *et al.*, in Peptides and Protein Phosphorylation, (ed.) Kemp, B.E. (CRC Press, Boca Raton, FL), pp. 172-187 (1990)], and synthetic peptides containing serine to threonine substitutions are relatively poor substrates for PKA phosphorylation [Zetterqvist, *et al.*, *supra*]. In the split-hybrid assay, however, it is unclear whether the mutation of threonine at position 133 disrupts the CREB-CBP interaction or if the mutant fails to become phosphorylated. Despite previous observations that serine residue at position 133 of mammalian CREB can be phosphorylated by a variety of protein kinases other than PKA, for example calcium/calmodulin-dependent protein kinase II and IV, protein kinase C, and a nerve growth factor (NGF)-activated CREB kinase [Sheng, *et al.*, *Neuron* **4**:571-582 (1990); Sheng, *et al.*, *Science* **252**:1427-1430 (1991); Xie and Rothstein, *J. Immunol.* **154**:1717-1723 (1995); Ginty, *et al.*, *Cell* **77**:1-20 (1994)], it is not known which, if any, of these particular protein kinases are able to phosphorylate CREB at the serine at position 133 in yeast. The requirement for integrity of the entire RRXSX amino acid sequence, however, suggests that PKA is a reasonable candidate.

The second category of mutations were identified adjacent the PKA phosphorylation motif. Amino acids isoleucine at position 137 and leucine at position 138 have previously been suggested to be important for hydrophobic interactions of CREB with CBP [Parker, *et al.*, *Mol. Cell. Biol.* **16**, 694-703 (1996)]. In this study, most of the mutations at position 137 and

138 converted these hydrophobic residues to polar amino acids. Thus, another possibility for the failure of these mutants to bind to CBP is that changes at these positions affect protein folding. Similarly, the mutation at position 141 substituted a polar residue for the wild-type hydrophobic leucine, and this mutation also has the potential to affect protein folding.

Substitution of the isoleucine at position 137 with a hydrophobic phenylalanine residue was found to disrupt the interaction between CREB and CBP as well. This result could have been the result of a detrimental effect on folding because of the steric hindrance associated with the comparatively larger size of phenylalanine. Alternatively, the proposed hydrophobic interactions between CREB and CBP are somewhat specific. Structural studies will be directed to definitively determine how these mutations affect binding.

Perhaps most surprising was the finding that critical mutations were restricted to a small region in the KID sequence, even though the relatively low affinity of phosphorylated CREB and CBP, determined to be between 250 and 400 nM by fluorescence anisotropy measurements [Kwok, *et al.*, *Nature* **370**, 223-226 (1994)], is consistent with a restricted protein binding domain. The capability of the split-hybrid system to screen for a limited number of CREB mutants suggests that the system is highly specific, and thus, should be useful to identify mutations which disrupt interacts between other pairs of binding proteins.

Example 4
Tax/SRF Binding Interaction

To further investigate the feasibility of using the split-hybrid system to study protein-protein interactions, a pair of well characterized interacting proteins, SRF and Tax, was tested. Previous studies indicated that SRF and Tax interact in a standard yeast two-hybrid system suggesting that the proteins may be utilized in the split hybrid assay. Plasmid pLexA-SRF, containing a human SRF cDNA fused to the LexA DNA binding domain, was transformed into strain YI584 along with either pVP16-Tax or pVP16 alone. As with the pLexA-VP16 transformation, the yeast strains co-expressing LexA-SRF and VP16-Tax failed to yield any colonies on medium lacking histidine. In contrast, when LexA-SRF was co-transformed with a vector encoding the VP16 activation domain alone, yeast growth occurred on medium lacking histidine, suggesting that TetR expression was not activated. These results demonstrated that a protein-protein interaction in the split-hybrid system can effectively prevent yeast growth and further indicated the utility of the assay for the study of various protein/protein interactions.

Example 5
Casein Kinase Binding Assays

Hrr25

In another example of use of the split hybrid assay to examine protein/protein interactions, Hrr25, a yeast casein kinase isoform, or human casein kinase I isoform δ, was employed in the assay with a known binding partner protein.

Previous work using the two hybrid assay had identified three genes encoding proteins which interact with the yeast casein kinase isoform Hrr25. Proteins encoded by the genes were designated TIH1, TIH2, and TIH3. The Hrr25 expression construct which was generated for use in the two hybrid assay was used in combination with the individual TIH encoding constructs in the split hybrid assay to determine if interaction between the

binding partners would decrease growth of assay yeast cells on media lacking histidine. Construction of the Hrr25 expression plasmid and isolation of plasmids encoding TIH proteins is discussed below.

In order to identify genes encoding proteins that interact with *S. cerevisiae* HRR25 CKI protein kinase, a plasmid library encoding fusions between the yeast GAL4 activation domain and *S. cerevisiae* genomic fragments ("prey" components) was screened for interaction with a DNA binding domain hybrid that contained the *E. coli lexA* gene fused to HRR25 ("bait" component). The fusions were constructed in plasmid pBTM116 which contains the yeast TRP1 gene, a 2μ origin of replication, and a yeast ADHI promoter driving expression of the *E. coli lex*A protein containing a DNA binding domain (amino acids 1 to 202).

Plasmid pBTM116::HRR25 encoding the *lexA*::*HRR25* fusion protein was constructed in several steps. The DNA sequence encoding the initiating methionine and second amino acid of HRR25 was changed to a *Sma*I restriction site by site-directed mutagenesis using a MutaGene mutagenesis kit from BioRad (Richmond, California). The DNA sequence of HRR25 is set out in SEQ ID NO: 39. The oligonucleotide used for the mutagenesis is set forth below, wherein the *Sma*I site is underlined.

5'-CCTACTCTTAGG<u>CCCGGG</u>TCTTTTTAATGTATCC-3'
(SEQ ID NO: 37)

After digestion with *Sma*I, the resulting altered HRR25 gene was ligated into plasmid pBTM116 at the *Sma*I site to create the lexA::HRR25 fusion construct.

Interactions between bait and prey fusion proteins were detected in yeast reporter strain CTY10-5d (genotype=*MATa ade2 trp1-901 leu2-3,112 his 3-200 gal4 gal80 URA3::lexA op-lacZ.*) [Luban, *et al., Cell 73*:1067-1078 (1993)] carrying a *lex*A binding site that directs transcription of

lacZ. Strain CTY10-5d was first transformed with plasmid pBTM116::HRR25 by lithium acetate-mediated transformation [Ito, *et al., J.Bacteriol. 153*:163-168 (1983)]. The resulting transformants were then transformed with a prey yeast genomic library prepared as GAL4 fusions in the plasmid pGAD [Chien, *et al., Proc.Natl.Acad.Sci (USA) 21*:9578-9582 (1991)] in order to screen the expressed proteins from the library for interaction with HRR25. A total of 500,000 double transformants were assayed for β-galactosidase expression by replica plating onto nitrocellulose filters, lysing the replicated colonies by quick-freezing the filters in liquid nitrogen, and incubating the lysed colonies with the blue chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal). β-galactosidase activity was measured using Z buffer (0.06 M Na_2HPO_4, 0.04 M NaH_2PO_4, 0.01 M KCl, 0.001 M $MgSO_4$, 0.05 M β-mercaptoethanol) containing X-gal at a concentration of 0.002% [Guarente, *Meth. Enzymol. 101*:181-191 (1983)]. Reactions were terminated by floating the filters on 1M Na_2CO_3 and positive colonies were identified by their dark blue color.

Library fusion plasmids (prey constructs) that conferred blue color to the reporter strain co-dependent upon the presence of the HRR25/DNA binding domain fusion protein partner (bait construct) were identified. The sequence adjacent to the fusion site in each library plasmid was determined by extending DNA sequence from the GAL4 region. The sequencing primer utilized is set forth below.

5'-GGAATCACTACAGGGATG-3' (SEQ ID NO: 38)

DNA sequence was obtained using a Sequenase version II kit (US Biochemicals, Cleveland, Ohio) or by automated DNA sequencing with an ABI373A sequencer (Applied Biosystems, Foster City, California).

Four library clones were identified and the proteins they encoded are designated herein as TIH proteins 1 through 4 for Targets Interacting with

HRR25-like protein kinase isoforms. The TIH1 portion of the TIH1 clone insert corresponds to nucleotides 1528 to 2580 of SEQ ID NO: 40; the TIH2 portion of the TIH2 clone insert corresponds to nucleotides 2611 to 4053 of SEQ ID NO: 41; and the TIH3 portion of the TIH3 clone insert corresponds to nucleotides 248 to 696 of SEQ ID NO: 42. Based on DNA sequence analysis of the TIH genes, it was determined that TIH1 and TIH3 were novel sequences that were not representative of any protein motif present in the GenBank database (July 8, 1993). TIH2 sequences were identified in the database as similar to a yeast open reading frame having no identified function. (GenBank Accession No. Z23261, open reading frame YBL0506)

When the various TIH proteins were used in the split hybrid assay in combination with Hrr25, it was observed that Hrr25/TIH3 binding, previously determined to be weaker than Hrr25/TIH2 or Hrr25/TIH1 interactions, produced the lowest level of growth in the transformed yeast strain.

CKIδ

In order to isolate cDNAs which encode proteins that interact with CKIδ, the two hybrid assay was performed using a LexA-CKIδ fusion protein as bait. The coding region of CKIδ was subcloned into a *Bam*HI site of pBTM116 and transformed into a yeast strain designated CKIδ/L40 (MAT a his3 Δ200 trp1-901 leu2-3 112 ade2 LYS::$(lexAop)_4$HIS3 URA3::$(lexAop)_8$-lcZ GAL 4). CKIδ/L40 was subjected to a large scale transformation with a cDNA library made from mouse embryos staged at days 9.5 and 10.5. Approximately 40 million transformants were obtained. Eighty-eight million were plated onto selective media lacking leucine, tryptophan and histidine. The ability of yeast transformants to grow in the absence of histidine suggested that there was an interaction between CKIδ and some library protein.

In a second screening, interaction of the two proteins was

assayed by the ability of the interaction to activate transcription of β-galactosidase. Colonies that turned blue in the presence of X-gal were streaked onto media lacking leucine, tryptophan and histidine, grown up in liquid culture and pooled for isolation of total DNA. Isolated DNA was used to transform *E. coli* strain 600 which lacks the ability to grow on media lacking leucine. Colonies that grew were used for plasmid preparation and three classes of cDNA were identified. One class was closely related to a *Drosophila* transcription factor dCREBa.

When CKIδ/CREB interaction was examined in the split hybrid assay, cells were shown to grow on media containing histidine, but in the absence of histidine, growth was inhibited. Addition of small amounts of tetracycline to the cell culture restored the cell's ability to grow, suggesting that the interaction between CKIδ and CREBa was very weak.

Example 6
AKAP 79 Binding Assays

Expression Plasmid Utilized

In still another example of use of the split hybrid assay to examine protein/protein interactions, an anchoring protein for the cAMP dependent protein kinase, AKAP 79, was utilized separately with binding partner proteins including the cAMP protein kinase regulatory subunit type I (RI), the cAMP dependent protein kinase regulatory subunit type II (RII) or calcineurin (CaN). Plasmids used in the assay were constructed as described below.

A 1.3 kb *Nco*I/*Bam*HI fragment containing the coding region of AKAP 79 was isolated from a pET11d backbone and ligated into plasmid pAS1. Plasmid pAS1 is a 2 micron based plasmid with an ADH promoter linked to the Gal4 DNA binding subunit [amino acids 1-147 as described in Keegan et al., *Science*, 231:699-704 (1986)], followed by a hemagglutin (HA) tag, polyclonal site and an ADH terminator. The expressed protein was

therefore a fusion between AKAP 79 and the DNA binding domain of Gal4.

Plasmids encoding RI, RII or CaN were isolated from a pACT murine T cell library in a standard two hybrid assay using the AKAP 79 expression construct described above. Plasmid pACT is a leu2, 2 micron based plasmid containing an ADH promoter and terminator with the Gal4 transcription activation domain II [amino acids 768-881 as described in Ma and Ptashne, *Cell*, 48:847-853 (1987)], followed by a multiple cloning site. RI, RII and CaN encoding plasmids were isolated as described below.

A 500 ml SC-Trp yeast cell culture (OD_{600} = 0.6-0.8) was harvested, washed with 100 ml distilled water, and repelleted. The pellet was brought up in 50 ml LiSORB (100 mM lithium acetate, 10 mM Tris pH8, 1 mM EDTA pH8, and 1 M Sorbitol), transferred to a 1 liter flask and shaken at 220 rpm during an incubation of 30 minutes at 30°C. The cells were pelleted, resuspended in 625 μl LiSORB, and held on ice while preparing the DNA.

The DNA was prepared for transformation by boiling 400 μl 10 mg/ml salmon sperm DNA for 10 minutes after which 500 μl LiSORB was added and the solution allowed to slowly cool to room temperature. DNA from a Mu T cell library was added (40-50 μg) from a 1 mg/ml stock. The iced yeast cell culture was dispensed into 10 Eppendorf tubes with 120 μl of prepared DNA. The tubes were incubated at 30°C with shaking at 220 RPM. After 30 minutes, 900 μl of 40% PEG_{3350} in 100 mM Li acetate, 10 mM Tris, pH 8, and 1 mM EDTA, pH 8, was mixed with each culture and incubation continued for an additional 30 minutes. The samples were pooled and a small aliquot (5 μl) was removed to test for transformation efficiency and plated on SC-Leu-Trp plates. The remainder of the cells were added to 100 ml SC-Leu-Trp-His media and grown for one hour at 30°C with shaking at 220 RPMS. Harvested cells were resuspended in 5.5 ml SC-Leu-Trp-His containing 50 mM 3AT (3-amino triazole) media and 300 μl aliquots plated on 150 mm SC-Leu-Trp-His also containing 50mM 3AT. Cell were left to

grow for one week at 30°C.

After four days, titer plates were counted and $1.1x10^5$ colonies were screened. Large scale β-gal assays were performed on library plates and ten positive clones were isolated for single colonies. One of these colonies grew substantially larger than the rest, and was termed clone 11.1. Sequence from clone 11.1 revealed an open reading frame 487 aa long which was correctly fused to the Gal-4 activation domain of pACT. The NIH sequence database was searched and the sequence was found to be closely homologous to the human calmodulin dependent protein phosphatase, calcineurin.

Additional screening using pACT Mu T-cell library DNA and the pASI AKAP 79 bait strain was performed in order to identify other AKAP 79 binding proteins by the protocol described above. Results from screening approximately 211,000 colonies gave one positive clone designated pACT 2-1. Sequencing and a subsequent data base search indicated that the clone had 91% identity with rat type 1α regulatory subunit of protein kinase A (RI).

The library was rescreened using the same AKAP 79 bait and fifteen positives were detected from approximately 520,000 transformants. Of these fifteen, eleven were found to be homologous to the rat regulatory subunit type I of PKA. Each of these isolates were fused to the 5' untranslated region of RI and remained open through the initiating methionine.

Split Hybrid Analysis

In split hybrid analysis of AKAP79 binding interactions, a plasmid was first constructed for expression of a LexA:AKAP 79 fusion protein. An AKAP 79 coding region was excised from pAS AKAP 79 as an *Nco*I/*Bam*HI fragment and inserted into pBTM116 previously digested with the same enzymes. The resulting plasmid was designated pBTM116-AKAP79.

Approximately 50,000 W303 yeast cells (strain YI665, see Table 1) in logarithmic growth were rinsed in media lacking histidine, suspended in 100 μl to 200 μl of the same media, and plated on agar lacking

histidine (to select for absence of protein/protein interaction) and also lacking leucine and tryptophan (to select for transformants bearing expression constructs encoding AKAP 79 and its binding partner). When RII was employed as the AKAP 79 binding partner, 2 to 4 μM tetracycline and 5 mM 3AT were required to prevent the transformed host from growing under conditions where the expressed proteins interacted.

Once conditions were established under which growth of the transformed host was eliminated, various candidate inhibitor compounds were separately added to the agar. It was presumed that if one of the candidate compounds was capable of disrupting AKAP 79 interaction with the binding partner protein, growth of the transformed host should be detectable in the vicinity of the compound on the agar. In the split hybrid assay wherein AKAP 79 and RII binding was examined, 2μl of a 30 mM stock solution of ICOS Compound 4273 in DMSO, 2 μl of a 10 mM stock solution of ICOS Compound 1062 in DMSO, and 2 μl DMSO alone (as a negative control) were spotted on to the plate which was incubated at 30°C for four to five days. For ICOS Compound 4273 a ring of growth was detected.

In order to determine an IC_{50} for an inhibitor identified as described above, alternative methods may be used. In one method, the inhibitor compound is added to the agar over a range of concentrations. Ideally, the compound is diluted to the point that host cell growth is essentially not detectable.

In another method, a 96 well plate is used and the compounds of interest are serially diluted across one row of a 96 well plate, one compound per row. Media lacking histidine, tryptophan, and leucine is added (presuming that the expression plasmids encoding the binding partners also encode trp and leu proteins) along with the appropriately transformed host yeast strain. Tetracycline and 3AT are added at concentration previously determined to extinguish growth of the transformed host cell. After two to five days incubation at 30°C, the plate wells are read at approximately 600

nm using a plate reader. The concentration of inhibitor half way between zero and the lowest concentration that permits growth of the host cell to the level observed on media containing histidine is estimated to be IC_{50}.

A modification of this second method is particularly amenable for use in a high throughput screen of large numbers of candidate inhibitors. For example, rather than attempting to determine the IC_{50} for a previously identified inhibitor, separate candidate inhibitors are added to each well of a 96 well plate, preferably at more than one concentration, and host cell growth determined after several days incubation. Inhibitory activity of compounds identified in this manner is confirmed on an agar plate and the IC_{50} determined on 96 well plates, each assay as described above.

Example 7
General Application of The Split-Hybrid Screen

In order to examine general utility of the split hybrid system, various experiments were conducted with binding proteins known to interact. In addition, a number of control experiments were included in order to determine if the effects observed with the known binding partners were in fact due to protein/protein interaction.

A. Yeast Assay Strain Construction

Yeast transformants used in assays indicated below were derived from LYS2-deficient strains AMR69 (Mat a *his3 lys2 leu2 trp1*, URA3:LexA::LacZ) and AMR70 (Mat α *his3 lys2 trp1 leu2*, URA3:LexA::LacZ) [Hollenberg, et al., *Mol. Cell. Biol.* **15**, 3813-3822 (1995); Chien, *et al., Proc. Natl. Acad. Sci. (USA)* **88**:97578-9582 (1991); Fields and Song, *Nature* **340**:245-246 (1989)]. Yeast were grown in YEPD or selective minimal medium using standard conditions [Sherman, F., et al., Methods in Yeast Genetics, Cold Spring Harbor Lab., Cold Spring Harbor, NY (1986); *Methods in Enzymology*, Vol. 194 Guide to Yeast Genetics and

Molecular Biology. Eds. Christine and Fink]. Derivatives of both AMR69 and AMR70 strains lacking URA3 were first generated by streaking cells on synthetic media containing 5 mg/ml 5-fluoro-orotic acid (5FOA) *[Methods in Enzymology*, Vol. 194 Guide to Yeast Genetics and Molecular Biology. Eds. Christine and Fink]. Two URA3 deficient mutants were required due to the fact that these strains were subsequently mated. URA3-deficient colonies were confirmed by testing for uracil auxotrophy and deletion of the URA:LexA::LacZ locus was confirmed by an absence of β-galactosidase activity assayed by standard methods. The mutant strains selected were designated 69-4 and 70-1.

Targeted integration of pRS306/8xLexAop/TetR was carried out by transforming [Hollenberg, *et al.*, *Mol. Cell. Biol.* **15**, 3813-3822 (1995)] the 69-4 strain with plasmid linearized at a unique *Nco*I site. The reporter gene construct was constructed using parental plasmid pRS306 which encodes *URA3* as a selectable marker. Stably integrated plasmid thereby permitted selection on media lacking uracil. The positive uracil prototrophic strains were examined by Southern analysis to confirm insertion of the plasmid sequences.

Targeted integration of pRS303/2xtetop-LYS was carried out by transformation [Hollenberg, *et al.*, *supra*] of strain 70-1 with plasmid linearized at a unique *Hpa*I site. The resulting lysine prototrophic strains were examined by Southern analysis to confirm insertion of the plasmid DNA.

The AMR69 derivative strain (MAT α) containing the pRS303/2xtetop-LYS insertion was mated with the AMR70-derivative strain (MAT a) containing pRS306/8xLexAop/TetR and mated cells were selected on media lacking both lysine and uracil. Single colonies were grown up and tested for the ability to grow on media lacking histidine. The resulting strain was designated YI584. In instances where yeast strains were transformed with other reporter gene pair combinations, the strains were uniquely designated.

Yeast bearing integrated reporter gene constructs were

subsequently transformed [Hollenberg, *et al.*, *supra*] with plasmids encoding chimeric binding protein. Plasmids encoding the LexA DNA binding region were generally derived from parental plasmid pBTM116 which also encodes *TRP1* as a selectable marker. Plasmids encoding the VP16 transactivating domain were generally derived from parental plasmid pVP16 which also encodes *LEU2* as a selectable marker. Yeast cells which were successfully transformed with the four exogenous plasmids were therefore selected by an ability to grow on media lacking lysine, uracil, tryptophan, and leucine. Plasmids encoding various binding proteins were transformed into the yeast assay strain as indicated below.

B. Liquid Assay

After three days growth at 30°C on selection media as described above, a pool of colonies from each transformation was collected and diluted in 5 ml selective media. The mixture was vortexed and immediately sonicated for ten seconds. Cells in the resulting suspension were counted and seeded at 1000 cells/ml in selective media, 2 ml per 15 ml tube. Tetracycline, 3AT, and histidine were included as determined appropriate by the method described above. Each aliquot of cells was incubated with shaking for two days at 30°C and cell density measured at OD_{600}.

C. Characterization of the Assay

The utility of the split-hybrid assay was first determined using well characterized binding proteins and various controls.

In an initial study, YI584 cells were transformed with plasmids pLexA-VP16 and pLeu. While the expressed proteins from the two plasmids do not interact, pLexA-VP16 encodes a fusion protein containing the VP16 activation domain fused directly to LexA which contains a DNA binding domain. The chimeric LexA-VP16 protein is a strong transactivator for a promoter containing LexA operators. Plasmid pLeu is essentially a blank used

as a control co-transformation plasmid.

Yeast transformed with the LexA-VP16 plasmid were able to express TetR protein as indicated by gel shift analysis using a *tet* operator oligonucleotide. In addition, the cells were unable to grow on media in the absence of histidine. Combined, these observations suggested that overexpressed TetR protein was capable of binding to *tet* operators and preventing the expression of *HIS3*. The transformed yeast grew on plates containing histidine, further indicating that overexpression of TetR did not have a toxic effect on the assay cells.

The results were consistent with previous observations and supported the earlier suggestion that activation of TetR expression, either through a single transcription factor or association of individual transcription factor domains, is capable of preventing assay cell growth on media lacking histidine, presumably by eliminating *HIS3* production.

Example 8
Split-Hybrid Assay With Weakly Interacting Binding Proteins

Protein/protein interaction was examined in the split-hybrid assay to determine utility of the system using two fusion proteins known to interact weakly. In this instance, the binding proteins were a 283 amino acid fragment of a cAMP regulatory binding protein (CREB283) fused to LexA and a fragment of the CREB binding protein consisting of the CREB binding domain (CBD) fused to VP16.

In this assay, yeast strain YI584 described above was employed and transformation carried out as previously described. In a first assay, plasmids pLexA-CREB and pVP16-CBD were transformed into the cells and cell growth was observed in the absence of histidine in the media. Expression of the fusion proteins was confirmed by Western blotting. Attempts to decrease cell growth by titration with 3AT were unsuccessful in that the concentration of 3AT required to reduce growth in cells transformed with

pLexA-CREB and pVP16-CBD also eliminated growth in cells transformed with pLexA-CREB and the control plasmid pVP16.

In light of these results, two alternative approaches were taken in order to permit study of binding proteins wherein the interaction is relatively weak. Under the assumption that the system was failing at the level of TetR transcription, alternative approaches were taken in attempts to amplify the TetR effect on expression of *HIS3* gene. To achieve this end, assay cells were transformed with reporter constructs which encoded multiple *tet* operator sequences upstream from the *HIS3* gene. In the second approach, the *HIS3* promoter used to drive expression of the TetR gene was replaced with the stronger alcohol dehydrogenase (ADH) promoter.

In YI596 cells wherein the ADH promoter replaced the *HIS3* promoter to drive TetR expression, transformation with plasmids pLexA-CREB and pVP16-CBD showed substantially decreased growth on his^- media as compared to that in assay strain YI592 wherein the *HIS3* promoter was used to drive TetR expression. However, in cells transformed with plasmids pLexA-CREB 341-M1 and pVP16-CBD, no decrease in assay cell growth was detected on media lacking histidine. These results indicate that incorporation of the ADH promoter to drive TetR expression may be more useful in studies involving binding proteins that have low affinity.

When assay strains were utilized which incorporated plasmids wherein expression of the *HIS3* gene was driven by multiple copies of the *tet* operator, transformed cell lines did not grow well enough to indicate potential utility in subsequent assays.

Example 9
General Assay Methods

A. "Fine Tuning"

In instances where either of the test fusion proteins possesses intrinsic capacity for transcriptional activation, TetR will be expressed and growth of the assay strain media lacking histidine will be depressed proportional to the level of TetR expression. In order to restore growth of these cells to approximately the level observed on media containing histidine, the initially transformed assay yeast strains are grown in the presence of increasing concentrations of tetracycline which binds to the TetR gene product and prevents TetR binding to the *tet* operator. Precise titration of expressed TetR with tetracycline, only to the point that growth of the assay strain is restored to the level detected in the presence of histidine, permits detection of subsequent decreased growth of the assay strain following increased TetR expression resulting from interaction of the test binding proteins. The empirically determined tetracycline concentration is therefore employed to increase "signal-to-noise" ratios under assay conditions.

After an appropriate tetracycline concentration has been determined for each of the candidate assay strains, the cells are transformed with the second plasmid encoding the second fusion binding protein. As before, growth of each candidate assay strain is examined on media in the presence and absence of histidine. A desirable yeast assay strain is chosen which shows vigorous growth in the presence of histidine and negligible growth on media lacking histidine (indicative of the expected protein/protein interaction and resultant decreased expression of *HIS3*).

In instances where binding between the two test proteins is comparatively weak, TetR expression may not be sufficiently increased to abolish *HIS3* expression and cells expressing the resultant low levels of *HIS3* will still grow on media which lacks histidine. Cells which show this low level of viability are grown in the presence of increasing concentrations of 3-

aminotriazole (3AT), a competitive inhibitor in the histidine synthesis pathway, in order to reduce cell growth to negligible levels when plated on media lacking histidine. As with titration of TetR with tetracycline, addition of 3AT to the media is designed to increase the signal-to-noise ratio by providing significant changes in growth in the presence and absence of histidine in the media.

In a practical application of the methods for fine tuning, binding between CREB and the CREB binding protein (CBP) is illustrative. Growth of the yeast strain YI584 transformed with pLexA-CBD, encoding the CREB binding domain (CBD) of CBP, and pVP16-CREB or pLexA-CBD and the control plasmid pVP16 was substantially decreased and virtually indistinguishable growth rates were detected in both instances on media lacking histidine. This observation indicated that the LexA-CBD protein product possessed sufficient transactivating capacity to eliminate HIS3 production. In order to distinguish growth differences between assay cells transformed with either pVP16 and pVP16-CREB, increasing amounts of tetracycline were added to the media lacking histidine.

In both transformants, tetracycline was able to relieve growth repression in a dose dependent manner, and at increasing concentrations of tetracycline, the difference in growth between the two colonies was increasingly magnified, with the most distinct growth difference observed following addition of tetracycline at 10 μg/ml. Addition of tetracycline was therefore able to overcome the intrinsic transactivating capability of the LexA-CBD fusion protein.

Because the ultimate use of the split-hybrid system is for structure-function studies, mutagenesis studies, drug identification and library screens, it is important to minimize background growth that might be confused with disrupted protein-protein associations. This can be accomplished by the addition of 3AT, a competitive inhibitor of the *HIS3* gene product. For instance, in the presence of 10 μg/ml of tetracycline, the yeast strain

transformed with pLexA-CBD and pVP16-CREB still conferred approximately 12% growth of that observed in the presence of his^+ media. To diminish this background, increasing concentrations of 3AT were added to the media in the presence of 10 μg/ml of tetracycline. At the 3AT concentration of 0.25 mM, the growth of the yeast strain expressing LexA-CBD and VP16-CREB was below 5%, while the growth of the control strain was still maintained at 70% of control levels. These results indicate that split-hybrid system can be modulated by 3AT in addition to tetracycline in order to effectively increase the signal-to-noise ratio.

B. Preparation of yeast extracts

In order to assess the utility of various plasmids to function in the split-hybrid assay, a number of control experiments can be employed which lend insight into expression of a desired protein from the transformed plasmid. For example, standard immunological methodologies, *i.e.*, immunoprecipitation, ELISA, *etc.*, can be used to determine to the extent to which a desired protein is expressed. Similarly, a variation of the gel shift assay (discussed immediately hereafter) can be used to determine both if a protein is expressed and if the expressed protein is capable of DNA binding. In each of these control assays, a yeast extract is required which can be prepared as follows.

Extracts were prepared as described by Uppaluri and Towle [*Mol. Cell. Biol.* **15**:1499-1512 (1995)] and were used for electrophoretic mobility shift assays as discussed below. The yeast cells transformed with pLexA-VP16 were grown in 100 ml of selective synthetic medium lacking uracil, tryptophan, and lysine to a density of A_{600} = 1. Cells were harvested and washed with 5 ml of EB (containing 0.2 M Tris-HCl, pH 8.0, 400 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$, 1 mM EDTA, 10% glycerol, and 7 mM β-mercaptoethanol). Cells were transferred to microcentrifuge tubes and collected by centrifugation. After resuspending in 200 μl EB containing 1

mM phenylmethylsulfonyl fluoride (PMSF), 1μg/ml leupeptin, and 1μg/ml pepstatin, a one-half volume of glass beads was added. The suspension was frozen in a -80°C freezer for 1 hour and thawed on ice. Thawed cells were vortexed at 4°C for 20 minutes, after which an additional 100 μl EB was added, and cells were left on ice for 30 minutes. The suspension was centrifuged for 5 minutes, the supernatant was transferred to a new tube which was centrifuged for 1 hour in a microcentrifuge. The supernatant was then made to 40% with $(NH_4)_2SO_4$ and gently rocked for 30 minutes. After a 10 minute centrifugation, the pellet was resuspended in 300 μl of 10 mM HEPES, pH 8.0, 5 mM EDTA, 7 mM β-mercaptoethanol, 1 mM PMSF, 1 μg/ml leupeptin, and 1 μg/ml pepstatin, and 20% glycerol. The resulting suspension was dialyzed against the same buffer, and aliquots were stored at -80°C.

C. Electrophoretic mobility shift assays

Electrophoretic mobility shift assays were performed as described by Shih and Towle [*J. Biol. Chem.* **267**:13222-13228 (1992)]. Double-stranded *tet* operator oligonucleotides were prepared by combining equivalent amounts of complementary single-stranded DNA (SEQ ID NOS: 7 and 8) in a solution containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, and 50 mM $NaCl_2$, heating the mixture to 70°C for 10 minutes, and then cooling to room temperature. The annealed oligonucleotides were labeled by filling in overhanging 5´ ends using the Klenow fragment of *E. coli* DNA polymerase I with [α-^{32}P]dCTP. Binding reactions were carried out in 20 μl containing 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 5% glycerol, and 2 mg of poly[d(I·C)]. A typical reaction contained 20,000 cpm (0.5-1 ng) of end-labeled DNA with 3-5 μg of yeast extract. Following incubation at 22°C for 30 minutes, samples were separated on a 4.5% nondenaturing polyacrylamide gel containing 50 mM Tris, 384 mM glycine, and 2 mM EDTA, pH 8.3. For competition binding experiments, the

conditions were exactly as above except that specific and nonspecific competitor DNAs were included in the binding mixture before the yeast extract was added. The concentration of tetracycline, a competitive inhibitor of TetR/*tet* operator binding, was 1 μM when utilized.

Example 10
Application of the Split-Hybrid Assay to Identify Agents That Prevent Receptor Desensitization and Drug Tachyphylaxis

Over half of the drugs that are used clinically affect the function of seven transmembrane receptors. Although many of the characteristics of these receptors are distinct, two general features appear to be conserved. One is the ability to signal through dissociation of heterotrimeric G proteins. The second is the capacity to lose responsiveness to ligand binding in a process termed desensitization which is mediated by receptor phosphorylation and the subsequent binding of factors that recognize the phosphorylated state of the receptor which prevents continued signaling. Desensitization results in an intrinsic limitation to drug action imposed by the action of the drug itself, *i.e.*, activation of a receptor by a hormone or drug initiates mechanisms that prevent subsequent responses to repeated administration of the same agent. The coupled mechanisms of activation and deactivation together have been termed "homologous desensitization," while the inability of a drug to maintain its efficacy is known as "tachyphylaxis." Even though the mechanisms underlying homologous desensitization have been worked out in great detail over the past few years, there are currently no useful pharmacological approaches available that prevent the inactivation mechanism.

The potential clinical utility of agents that could prevent or modulate drug desensitization is enormous. Four examples where therapy is limited by the inability of receptors to maintain responsiveness to drugs include: (i) asthma wherein desensitization of airway adrenergic receptors

renders epinephrine treatment ineffective after a period of hours; (ii) congestive heart failure wherein desensitization of adrenergic and VIP receptors, coupled with an elevation of the β adrenergic receptor kinase (βARK), prevents the inotropic effects of endogenous regulatory hormones; (iii) Parkinson's disease, wherein dopamine receptor desensitization limits the usefulness of agents like L-Dopa; and (iv) chronic pain wherein tolerance results from opiate receptor desensitization. Indeed, it is difficult to conceive of a pharmacological modality in use today that is not limited in its effectiveness by the phenomenon of desensitization.

The biochemical basis for G protein-coupled receptor desensitization involves three classes of proteins including arrestins, kinases and G-proteins, all of which have been cloned [Lefkowitz, *Nature Biotechnology* **14**:283-286 (1996)]. Following activation of a seven transmembrane receptor, a region is phosphorylated by one or more G protein-coupled receptor kinases (known as GRKs 1-6). For example, in the β-adrenergic receptor (βAR) and rhodopsin, the cytoplasmic tail is phosphorylated [Premont, *et al., J. biol. Chem.* **269**:6832-6841 (1994); Freedman, *et al., J. Biol. Chem.* **270**:17953-17961 (1995); Palczewski,*et al., J. Biol. Chem.* **266**:12949-12955 (1991); Palczewski, *et al., J. Biol. Chem.* **270**:15294-15298 (1995)] while in the m2 muscarinic receptor, the third cytoplasmic loop is phosphorylated [Nakata, *et al., Eur. J. Biochem.* **220**:29-36 (1994)]. The best characterized members of the family of G protein receptor kinases are the βAR kinase (βARK) and rhodopsin kinase which are both membrane-associated. While rhodopsin kinase contains an intrinsic membrane targeting signal [Inglese, *et al., Nature* **359**:147-150 (1992)], βARK appears to be targeted to the membrane by association with G protein βγ subunits [Pitcher, *et al., Science* **257**:1264-1267 (1992); Inglese, *et al., Nature* **359**:147-150 (1992)]. Once the substrate receptor for each kinase is activated, presumably by ligand binding, the kinase associates and phosphorylates serine and threonine residues on the receptor. The phosphorylated receptor then becomes a binding target for one or more

other proteins. In the case of βAR, for example, phosphorylation allows binding of arresting which prevents association with G proteins and promotes receptor sequestration and desensitization. Using the βAR as an exemplary desensitization model, it becomes apparent that multiple steps in the pathway appear to provide potential points of regulation each of which is amenable to the split-hybrid screen to identify molecules that can block the overall desensitization pathway. Specifically in the case of βAR, the split hybrid system can be used to identify small molecules that: (i) prevent interaction between βARK and the G protein β subunit; (ii) inhibit βARK activity; and (iii) disrupt the βARK:arresting complex.

A. Plasmid Constructions

The study of G-protein receptor kinases in the split-hybrid system involves three or more recombinant proteins or two or more recombinant proteins and a recombinant peptide library. In the split-hybrid system discussed above, two yeast primary expression plasmids are employed: pBTM116 [Bartel *et al., Cellular Interactions in Development: a Practical Approach*, (ed) Hartley, IRL Press, Oxford, pp. 153-179 (1993)], which encodes the LexA-fusion protein and the *TRP1* selectable marker, and pVP16 [Hollenberg *et al., Mol. Cell. Biol.*, 15:3813-3822 (1995)], which encodes the VP16-fusion protein and the *LEU2* selectable marker. In order to study interactions involving more than two recombinant proteins in the split-hybrid system, however, additional selectable markers are employed. Construction of additional yeast expression plasmids which are used to examine interactions between more than two binding proteins is discussed below.

1. Plasmid pDRM

A DNA fragment comprising the ADH promoter and LexA sites, the TetR encoding gene, the nuclear localization signal, and the ADH terminator sequence are removed from pRS306/4xLexAop/ADH::TetR with *Sac*I, blunt-ended, and digested with *Sal*I. The fragment is isolated and ligated into pRS303/2xtetop-LYS2 which has previously been digested with *Not*I, blunt-ended, and digested with *Sal*I. The resulting plasmid, designated pDRM, is integrated into the *LYS2* locus in the yeast genome as described above, and the resulting strain designated YIDRM. Placing the repressor gene and selectable marker reporter gene in the *LYS2* locus allows *ERA3* to be used a selectable marker.

2. Plasmid pRSURA3

A modified version of the pRS306 vector [Sikorski *et al.*, *Genetics*, 122:19-27 (1989)] containing the *URA3* selectable marker gene is also used to encode additional recombinant proteins in the split-hybrid system. The plasmid, pRS426, has the 2 micron origin of replication inserted into a unique *Aat*I site of pRS306. Plasmid pRS426 is further modified in the following manner:

(i) The ADH promoter sequence is amplified by PCR from BTM116 using primers which incorporate into the amplification product the DNA sequence encoding the SV40 large T antigen nuclear localization signal (NLS) and an initiating ATG sequence 3' to the ADH promoter. The ADH promoter/NLS/ATG sequence is inserted into the polylinker of pRS426.

(ii) The ADH terminator sequence is amplified by PCR from BTM116 using primers which incorporate into the product a DNA sequence encoding an antibody tag, for example, FLAG, hemagglutinin protein (HA), or thioredoxin (Thio) (FLAG, HA, and Thio antibodies are available through Santa Cruz Biotechnology, Santa Cruz, CA) and DNA sequences encoding stop codons in all three frames to the 5' end of the ADH terminator sequence.

The antibody tag/stop codon/ADH terminator sequence is inserted into the polylinker of pRS426.

3. Plasmid pRSADE2

PCR is used to engineer unique restriction sites, including for example, *Bgl*II, *Eco47*III, *Mlu*I, *Nhe*I, and *Sph*I, immediately adjacent the 5' and 3' ends of the *URA3* cassette in pRSURA3. The *URA3* cassette is digested from pRSURA3 and replaced with the *ADE2* cassette which is amplified by PCR.

4. Plasmid pBTM116/AD4

A fragment containing the ADH promoter, polylinker, and ADH terminator is digested from pAD4 [Young *et al., Proc. Nat'l. Acad. Sci. (USA), 86:*7989-7993 (1989)] with *BamH*I, blunt-ended and inserted into the blunt-ended *Pvu*I site of BTM116 as described [Keegan *et al., Oncogene, 12:*1537-1544 (1996)], and the resulting vector designated pBTM116/AD4. PCR is also used to engineer a nuclear localization signal 3' of the ADH promoter as described above. This vector contains the *TRP1* selectable marker and can encode two recombinant proteins: (i) a LexA-fusion protein and (ii) a protein expressed from the pAD4 region of the vector.

B. βARK and G Protein β Subunit Binding

In a first application of the split hybrid assay, disruption of binding between the carboxy-terminal domain of βARK, containing the pleckstrin homology (PH) domain, and the G protein β subunit ($G\beta_2$) is examined. Previous work indicates that the PH domain of βARK interacts directly with the βγ subunits of G proteins [Pitcher, J.A., *et al. Science* **257**:1264-1267 (1992) and Touhara, K. *et al., J.Biol.Chem.* **269**:10217-10220 (1994)]. Consistent with this observation is work by Pumiglia, *et al.* [Pumiglia, K.M., *et al., J.Biol.Chem.* **270**:14251-14254 (1995)] which

indicates that $G\beta_2$ interacts with Raf1 in yeast and that the interaction is disrupted by βARK *in vitro*.

A DNA fragment containing the carboxy-terminal 222 amino acids (residues 467 to 689) of βARK1, which includes the PH domain, is amplified by PCR from bovine βARK1 [Pitcher *et al., Science, 257:*1264-1267 (1992)] and the gel-purified amplification product is inserted into pBTM116. The resulting plasmid is designated LexA-COOH-βARK. A DNA fragment containing the entire coding sequence of $G\beta_2$ [Fong *et al., Proc. Nat'l. Acad. Sci. (USA), 84:*3792-3796 (1987)] is PCR amplified from pGEM-11Zf(-)$G\beta_2$ [Inigez-Lluhi *et al., JBC, 267:*23409-23417 (1992)] and the gel-purified amplification product inserted into pVP16. The resulting plasmid is designated pVP16-$G\beta_2$. PCR is used in a similar manner to clone the carboxy-terminal domain of βARK into pVP16 and $G\beta_2$ into pBTM116.

βARK and $G\beta_2$ binding is first examined in the two-hybrid system to determine if expression of either binding partner as a fusion protein in yeast affects protein/protein interaction. Binding of the two proteins is then examined in the split hybrid assay in order to determine if protein/protein interaction is capable of abolishing growth of the assay yeast strain. As above, addition of tetracycline and/or 3-aminotriazole required to maximize the difference in growth in the presence and absence of histidine is empirically determined.

Split-hybrid yeast strains containing βARK and $G\beta_2$ subunits are used to screen libraries of small molecules. Several types of small molecule libraries can be examined in the split-hybrid assay, including for example, chemical libraries, libraries of products naturally produced by microorganisms, animals, plants and/or marine organisms, combinatorial, recombinatorial, peptidomimetic, multiparallel synthetic collection, protein, peptide and polypeptide libraries. A library of small peptides can be cloned into pRSURA3 as described [Yang *et al., Nuc. Acids Res., 23:*1152-1156 (1995) and Colas *et al., Nature, 380:*548-550)]. Peptides corresponding to the

carboxy-terminus of βARK or other GRKs which have previously been shown to block calcium channel desensitization in intact neurons, presumably by blocking βARK and $G\beta_2$ binding and subsequent trafficking of βARK to the cellular membrane [Diverse-Pierluissi, *et al., Neuron* **16**:579-585 (1996)] can be identified in such a screen. Further, it is important to show that the molecules identified through the split hybrid selection affect βARK:Gβ interaction as opposed to, for example, tetracycline analogues identified in the screen that would not be useful to specifically modulate βARK/$G\beta_2$ binding.

B. Identification of βARK Inhibitors

In a second approach, agents that directly inhibit βARK function are identified in a modification of the split-hybrid system. While identification of specific βARK inhibitors may be difficult, preliminary data from split hybrid assays using CREB/CBP binding partners indicates that the system can be used to identify serine kinase inhibitors. The serine kinase results also suggest several approaches can be employed in attempts to overcome potential problems in identifying βARK inhibitors.

Briefly, binding between the phosphorylated G-protein coupled receptor (P-GR) and arresting is examined first in the standard two hybrid assay, followed by identification of inhibitors of P-GR/arresting binding in the split hybrid assay. For these studies, fragments of three G protein-coupled receptors are examined: the carboxy-terminal tail of β_2AR and the third cytoplasmic loop of the m2 muscarinic receptor. A DNA fragment containing the carboxy-terminal tail of the β_2AR (amino acids 330 to 413) is PCR amplified [Kolbilka *et al., JBC, 262:*7321-7327 (1987)] and the gel purified product inserted into pBTM116/Ad4 to produce a LexA-β_2AR fusion gene. The resulting plasmid is designated pBTM-β_2AR/AD4. A DNA fragment containing the third cytoplasmic loop of the human m2 muscarinic receptor (nucleotides 268-324) is amplified from pGEX-I3m2 [Haga *et al., JBC,*

*269:*12594-12599 (1994)] by PCR and cloned into pBTM116/Ad4 creating a LexA-m2 fusion gene. The resulting plasmid is designated pBTM-m2/AD4. The entire bovine βARK1 coding sequence [Benovic *et al., Science, 246:*235-240 (1989)] is PCR amplified and cloned into the polylinker region originating from AD4 in pBTM-β_2AR/AD4 and pBTM-m2/AD4. The resulting plasmids are designated pBTM-β_2AR/AD4-βARK and pBTM-m2/AD4-βARK, respectively. PCR is used to amplify the DNA fragment containing bovine βarresting-1 (amino acids 1 to 437) [Lohse, *et al., Science, 248:*1547-1550 (1990)]. This fragment is inserted into pVP16 and is designated pVP16-βarresting-1. PCR is used to amplify the DNA fragment containing rat βarresting-2 (amino acids 1 to 428) [Attramadal, *et al., JBC, 267:*17882-17890 (1992)] which is inserted into pVP16 to give plasmid pVP16-βarresting-2. A PCR strategy is also used to clone arresting into the pBTM116/AD4-βARK plasmid and the βAR and m2 fragments into pVP16. As above, the yeast split-hybrid YIDRM strain is transformed with the P-GR-arresting along with peptide libraries (cloned into pRSURA3) or grown following transformation in the presence of combinatorial drug libraries.

Inhibitors identified in the split hybrid assay should effect disruption of protein/protein interaction either by: (i) inhibiting βARK phosphorylation of the receptor, thus preventing recognition of the receptor by arresting, or (ii) by physical disruption of binding between the receptor and arresting. Agents that allow yeast growth for trivial reasons, *i.e.*, tetracycline analogues, can be easily identified through use of simple controls.

A first potential problem to overcome in this study is that cytoplasmic βARK enzyme must be targeted to the substrate receptor and, once targeted, must phosphorylate the receptor at appropriate sites. In normal cells, βγ association serves to target βARK to the cell membrane; the β subunit binds to both the βARK PH domain and the isoprenylated γ subunit in association with the membrane. One possible means to encourage the necessary specific interactions is to target the binding components in the assay

by tagging the proteins with nuclear localization signals, *i.e.*, βARK, the receptor cytoplasmic tail, and arresting, to the nucleus. The plasmids proposed for the study of the P-GR-arresting interaction all contain nuclear localization signal sequences adjacent to recombinant gene sequence.

A second problem is somewhat more difficult to approach. The current model is that receptors must be activated by ligand binding before being phosphorylated by βARK, *i.e.*, targeting of βARK via βγ is not sufficient for receptor phosphorylation. There are two possible explanations for this requirement. The first is that phosphorylation sites on the receptor are masked in the absence of ligand and ligand binding causes a conformational change which "unmasks" the phosphorylation sites. If this is the case, a fragment of the receptor containing the immediate phosphorylation site may be used as the βARK target. However, although peptides representing portions of the βAR cytoplasmic tail can be phosphorylated by βARK, the K_m for the phosphorylation reaction is poor, suggesting that the kinase may require some other part of the receptor for binding and that the unmasking of this binding site by agonist is a critical step.

This problem is addressed in two ways. In the first, the m2 muscarinic receptor is used in place of the βAR in view of previous results which indicate that the m2 protein is a good substrate for βARK. The third cytoplasmic loop of the m2 receptor serves as both the binding site and phosphorylation site for kinase and which should allow use of a LexA/m2 receptor third cytoplasmic loop fusion gene as one component in the screening system.

An alternative approach is to artificially mimic the activated state of the receptor. Haga, *et al.* [*J. Biol. Chem.* **269**:12594-12599 (1994)] have shown that the activity of βARK can be stimulated *in vitro* in the presence of mastoporan, a bee venom peptide. Mastoporan is believed to mimic the cytoplasmic face of an activated receptor and has been shown to increase the affinity of βARK for a GST-m2 receptor fusion protein by over

four orders of magnitude. The same effect can be seen by using peptides representing the flanking regions of the m2 third cytoplasmic loop. Thus, mastoporan should also activate βARK in the two-hybrid yeast strains, allow phosphorylation of the receptor fusion protein, and promote interaction with arresting. If mastoparan is needed, oligonucleotides containing the coding and non-coding nucleotide sequences of the 14-mer peptide (INLKALAALAKKIL-NH_2, SEQ ID NO: 43) are annealed and ligated into prSADE2. The yeast split-hybrid strain YIDRM is transformed with pBTM-βAR (or m2)/AD4-βARK, pVP16-arresting, pRSADE2-mastoparan, and a pRSURA3-peptide library or combinatorial drug library.

Numerous modifications and variations in the invention as set forth in the above illustrative examples are expected to occur to those skilled in the art. Consequently only such limitations as appear in the appended claims should be placed on the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Hoekstra, Merl F.

(ii) TITLE OF INVENTION: Methods to Identify Compounds For Disrupting Protein/Protein Interactions

(iii) NUMBER OF SEQUENCES: 43

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
(B) STREET: 6300 Sears Tower, 233 South Wacker Drive
(C) CITY: Chicago
(D) STATE: Illinois
(E) COUNTRY: United States of America
(F) ZIP: 60606-6402

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME:
(B) REGISTRATION NUMBER:
(C) REFERENCE/DOCKET NUMBER: 27866/33424

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 312/474-6300
(B) TELEFAX: 312/474-0448
(C) TELEX: 25-3856

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTGGTGAGCG CTAGGAGTCA CTGCCAG 27

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TATACTCTAT CAATGATAGA GTAATTCATT ATGTGATAAT GCC 43

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATTACTCTAT CATTGATAGA GTATATAAAG TAATGTGATT TC 42

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AATTCTGCTA GCCTCTGCAA AGC 23

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGCACGCGTC GAAGAAATCA CATTACTTTA TATA 34

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCACGCGTA TACTAAAAAA TGAGCAGGCA AG 32

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGCGTACTCT ATCATTGATA GAGTA 25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGAGATAGT AACTATCTCA TGCGC 25

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGCGTACTCT ATCATTGATA GAGTCTAGAC TCTATCAATG ATAGAGTA 48

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCGACGCGTG CATGCCGTCT TCAAGAATTC CTCGAG 36

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCGACGCGTG CATGCCCACC GTACACGCCT ACTCGA 36

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CATGGCATGC AAAAAAAAAG AGTCATCCGC TAGG 34

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CATGGCATGC TTAGCGATTG GCATTATCAC AT 32

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TAATACGACT CACTATATAG GG 22

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TCTAGACTTT GCCTTCGTTT ATC 23

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGAAGGCAAA GATGTCTAGA TTAGATAAAA G 31

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CGCGGATCCG CTTTCTCTTC TTTTTTGGAG ACCCACTTTC ACATTTAAG 49

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AATTGCTCGA GTACTGTATG TACATACAGT AG 32

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AATTCTACTG TATGTACATA CAGTACTCGA GC 32

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CCGGAATTCT CGAGACATAT CCATATCTAA TC 32

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCGGAATTCA CTAATCGCAT TATCATC 27

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CATGCCATGG CCATGTCTAG ATTAGATAAA AG 32

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GCGAATTCGC CAGGGCAACA GAATGCCACT 30

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CGGGATCCTG GCTGGTTACC CAGGATGCCT TG 32

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CGCGGATCCG GATGACCATG GACTCTGGAG
30

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CGCGGATCCT TAATCTGACT TGTGGCAGTA
30

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CGCGGATCCC CATGACCATG GAATCTGGAG CC
32

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CGCGGATCCG TGCTGCTTCT TCAGCAGGCT G
31

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

ATGGTACCAG CGGCCGCTAG TCGTTTTACA ACGTCGTGAC 40

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ATGGTACCGC GGCCGCTTAT TTTTGACACC AGACCAAC
38

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CGGAGATCTA AAGAGACTTT TCTCCGGAAC TCAG
34

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CGGAGATCTT TACAGGAAGA CTGAACTGT
29

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CCACCGCGGC AGTGCCAACC CCGATTTAC
29

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CATCCGCGGT GGTGATGGCA GGGGCTGA
28

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGCTATCGAT ACGGCCCCCC CGACCGAT
28

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GCGTATCGAT CTACCCACCG TACTCGTC
28

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CCTACTCTTA GGCCCGGGTC TTTTTAATGT ATCC
34

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GGAATCACTA CAGGGATG
18

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1485 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

ATGGACTTAA GAGTAGGAAG GAAATTTCGT ATTGGCAGGA AGATTGGGAG TGGTTCCTTT 60

GGTGACATTT ACCACGGCAC GAACTTAATT AGTGGTGAAG AAGTAGCCAT CAAGCTGGAA 120

TCGATCAGGT CCAGACATCC TCAATTGGAC TATGAGTCCC GCGTCTACAG ATACTTAAGC 180

GGTGGTGTGG GAATCCCGTT CATCAGATGG TTTGGCAGAG AGGGTGAATA TAATGCTATG 240

GTCATCGATC TTCTAGGCCC ATCTTTGGAA GATTTATTCA ACTACTGTCA CAGAAGGTTC 300

TCCTTTAAGA CGGTTATCAT GCTGGCTTTG CAAATGTTTT GCCGTATTCA GTATATACAT 360

GGAAGGTCGT TCATTCATAG AGATATCAAA CCAGACAACT TTTTAATGGG GGTAGGACGC 420

CGTGGTAGCA CCGTTCATGT TATTGATTTC GGTCTATCAA AGAAATACCG AGATTTCAAC 480

ACACATCGTC ATATTCCTTA CAGGGAGAAC AAGTCCTTGA CAGGTACAGC TCGTTATGCA 540

AGTGTCAATA CGCATCTTGG AATAGAGCAA AGTAGAAGAG ATGACTTAGA ATCACTAGGT 600

TATGTCTTGA TCTATTTTTG TAAGGGTTCT TTGCCATGGC AGGGTTTGAA AGCAACCACC 660

AAGAAACAAA AGTATGATCG TATCATGGAA AAGAAATTAA ACGTTAGCGT GGAAACTCTA 720

TGTTCAGGTT TACCATTAGA GTTTCAAGAA TATATGGCTT ACTGTAAGAA TTTGAAATTC 780

GATGAGAAGC CAGATTATTT GTTCTTGGCA AGGCTGTTTA AAGATCTGAG TATTAAACTA 840

GAGTATCACA ACGACCACTT GTTCGATTGG ACAATGTTGC GTTACACAAA GGCGATGGTG 900

GAGAAGCAAA GGGACCTCCT CATCGAAAAA GGTGATTTGA ACGCAAATAG CAATGCAGCA 960

AGTGCAAGTA ACAGCACAGA CAACAAGTCT GAAACTTTCA ACAAGATTAA ACTGTTAGCC 1020

ATGAAGAAAT TCCCCACCCA TTTCCACTAT TACAAGAATG AAGACAAACA TAATCCTTCA 1080

CCAGAAGAGA TCAAACAACA AACTATCTTG AATAATAATG CAGCCTCTTC TTTACCAGAG 1140

GAATTATTGA ACGCACTAGA TAAAGGTATG GAAAACTTGA GACAACAGCA GCCGCAGCAG 1200

CAGGTCCAAA GTTCGCAGCC ACAACCACAG CCCCAACAGC TACAGCAGCA ACCAAATGGC 1260

CAAAGACCAA ATTATTATCC TGAACCGTTA CTACAGCAGC AACAAAGAGA TTCTCAGGAG 1320

CAACAGCAGC AAGTTCCGAT GGCTACAACC AGGGCTACTC AGTATCCCCC ACAAATAAAC 1380

AGCAATAATT TTAATACTAA TCAAGCATCT GTACCTCCAC AAATGAGATC TAATCCACAA 1440

CAGCCGCCTC AAGATAAACC AGCTGGCCAG TCAATTTGGT TGTAA 1485

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2625 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 796..2580

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
CATTTTCTTA ATTCTTTTAT GTGCTTTTAC TACTTTGTTT AGTTCAAAAC AATAGTCGTT      60

ATTCTTAGGT ACTATAGCAT AAGACAAGAA AAGAAAAATA AGGGACAAAT AACATTAGCA     120

GAAGTACGGT ATATTTTACT GTTACTTATA TACTTTCAAG AAGATGAGTT AAATCGGTAG     180

CCAGTGTAGA AAAATAATAA TAAGGGTCAT CGATCCTTCG CATTTTATTA TCCAATTAAA     240

GATACGAATC ACGGCAAACT ATATTCAAAG CTCATAGATA ATCGTCGTAA GGCTGACACT     300

GCAGAAGAAA AGTCATAATT TGAATACTAG CCGGTATGAA ACTGTGATTG ATTAACCTGG     360

GGTTACCTAA AGAGAACATA AGTAATACTC ATGACAGAAT CAAAACACAA TACAAAATTT     420

ATCCGAACCT CGGCCCGACT GCGGCTCGCC GGGAAAGGGG ACAACCGCTT CTATCCGTCG     480

ACTAACTTCA TCGGCCCAAT GGAAGCTATG ATATGGGGAT TTCCATTGAG CCGATAGCAA     540

TGTAGGGTAA TACTGTTGCG TATATAGTGA TAGTTATTGA ATTTTATTAC CCTGCGGGAA     600

TATTGAGACA TCACTAAGCA CGAATTTTAC GTCTGAGGAA AGTTGAATGA TGGCCAAATA     660

ACCAGGAAAA ACAAATATTG AATCCTTGTG AAGGATTCCA CAGTTGTTTA ATCCTCCTTA     720

AGCTCACTTA GTATCAATTG TCTAAATAAT ATTGCTTTGA ATCTGAAAAA AATAAAAGTA     780

CCTTCGCATT AGACA ATG TCA CTG CCG CTA CGA CAC GCA TTG GAG AAC GTT      831
                Met Ser Leu Pro Leu Arg His Ala Leu Glu Asn Val
                  1               5                  10

ACT TCT GTT GAT AGA ATT TTA GAG GAC TTA TTA GTA CGT TTT ATT ATA      879
Thr Ser Val Asp Arg Ile Leu Glu Asp Leu Leu Val Arg Phe Ile Ile
        15                  20                  25

AAT TGT CCG AAT GAA GAT TTA TCG AGT GTC GAG AGA GAG TTA TTT CAT      927
Asn Cys Pro Asn Glu Asp Leu Ser Ser Val Glu Arg Glu Leu Phe His
    30                  35                  40

TTT GAA GAA GCC TCA TGG TTT TAC ACG GAT TTC ATC AAA TTG ATG AAT      975
Phe Glu Glu Ala Ser Trp Phe Tyr Thr Asp Phe Ile Lys Leu Met Asn
45                  50                  55                  60

CCA ACT TTA CCC TCC CTA AAG ATT AAA TCA TTT GCT CAA TTG ATC ATA     1023
Pro Thr Leu Pro Ser Leu Lys Ile Lys Ser Phe Ala Gln Leu Ile Ile
                65                  70                  75

AAA CTA TGT CCT CTG GTT TGG AAA TGG GAC ATA AGA GTG GAT GAG GCA     1071
Lys Leu Cys Pro Leu Val Trp Lys Trp Asp Ile Arg Val Asp Glu Ala
            80                  85                  90

CTC CAG CAA TTC TCC AAG TAT AAG AAA AGT ATA CCG GTG AGG GGC GCT     1119
Leu Gln Gln Phe Ser Lys Tyr Lys Lys Ser Ile Pro Val Arg Gly Ala
        95                  100                 105

GCC ATA TTT AAC GAG AAC CTG AGT AAA ATT TTA TTG GTA CAG GGT ACT     1167
Ala Ile Phe Asn Glu Asn Leu Ser Lys Ile Leu Leu Val Gln Gly Thr
```

```
    110                 115                 120

GAA TCG GAT TCT TTG TCA TTC CCA AGG GGG AAG ATA TCT AAA GAT GAA      1215
Glu Ser Asp Ser Leu Ser Phe Pro Arg Gly Lys Ile Ser Lys Asp Glu
125                 130                 135                 140

AAT GAC ATA GAT TGT TGC ATT AGA GAA GTG AAA GAA GAA ATT GGT TTC      1263
Asn Asp Ile Asp Cys Cys Ile Arg Glu Val Lys Glu Glu Ile Gly Phe
                145                 150                 155

GAT TTG ACG GAC TAT ATT GAC GAC AAC CAA TTC ATT GAA AGA AAT ATT      1311
Asp Leu Thr Asp Tyr Ile Asp Asp Asn Gln Phe Ile Glu Arg Asn Ile
            160                 165                 170

CAA GGT AAA AAT TAC AAA ATA TTT TTG ATA TCT GGT GTT TCA GAA GTC      1359
Gln Gly Lys Asn Tyr Lys Ile Phe Leu Ile Ser Gly Val Ser Glu Val
        175                 180                 185

TTC AAT TTT AAA CCT CAA GTT AGA AAT GAA ATT GAT AAG ATA GAA TGG      1407
Phe Asn Phe Lys Pro Gln Val Arg Asn Glu Ile Asp Lys Ile Glu Trp
    190                 195                 200

TTC GAT TTT AAG AAA ATT TCT AAA ACA ATG TAC AAA TCA AAT ATC AAG      1455
Phe Asp Phe Lys Lys Ile Ser Lys Thr Met Tyr Lys Ser Asn Ile Lys
205                 210                 215                 220

TAT TAT CTG ATT AAT TCC ATG ATG AGA CCC TTA TCA ATG TGG TTA AGG      1503
Tyr Tyr Leu Ile Asn Ser Met Met Arg Pro Leu Ser Met Trp Leu Arg
                225                 230                 235

CAT CAG AGG CAA ATA AAA AAT GAA GAT CAA TTG AAA TCC TAT GCG GAA      1551
His Gln Arg Gln Ile Lys Asn Glu Asp Gln Leu Lys Ser Tyr Ala Glu
            240                 245                 250

GAA CAA TTG AAA TTG TTG TTG GGT ATC ACT AAG GAG GAG CAG ATT GAT      1599
Glu Gln Leu Lys Leu Leu Leu Gly Ile Thr Lys Glu Glu Gln Ile Asp
        255                 260                 265

CCC GGT AGA GAG TTG CTG AAT ATG TTA CAT ACT GCA GTG CAA GCT AAC      1647
Pro Gly Arg Glu Leu Leu Asn Met Leu His Thr Ala Val Gln Ala Asn
    270                 275                 280

AGT AAT AAT AAT GCG GTC TCC AAC GGA CAG GTA CCC TCG AGC CAA GAG      1695
Ser Asn Asn Asn Ala Val Ser Asn Gly Gln Val Pro Ser Ser Gln Glu
285                 290                 295                 300

CTT CAG CAT TTG AAA GAG CAA TCA GGA GAA CAC AAC CAA CAG AAG GAT      1743
Leu Gln His Leu Lys Glu Gln Ser Gly Glu His Asn Gln Gln Lys Asp
                305                 310                 315

CAG CAG TCA TCG TTT TCT TCT CAA CAA CAA CCT TCA ATA TTT CCA TCT      1791
Gln Gln Ser Ser Phe Ser Ser Gln Gln Gln Pro Ser Ile Phe Pro Ser
            320                 325                 330

CTT TCT GAA CCG TTT GCT AAC AAT AAG AAT GTT ATA CCA CCT ACT ATG      1839
Leu Ser Glu Pro Phe Ala Asn Asn Lys Asn Val Ile Pro Pro Thr Met
        335                 340                 345

CCA ATG GCT AAC GTA TTC ATG TCA AAT CCT CAA TTG TTT GCG ACA ATG      1887
Pro Met Ala Asn Val Phe Met Ser Asn Pro Gln Leu Phe Ala Thr Met
    350                 355                 360

AAT GGC CAG CCT TTT GCA CCT TTC CCA TTT ATG TTA CCA TTA ACT AAC      1935
```

```
Asn Gly Gln Pro Phe Ala Pro Phe Pro Phe Met Leu Pro Leu Thr Asn
365                 370                 375                 380

AAT AGT AAT AGC GCT AAC CCT ATT CCA ACT CCG GTC CCC CCT AAT TTT      1983
Asn Ser Asn Ser Ala Asn Pro Ile Pro Thr Pro Val Pro Pro Asn Phe
                385                 390                 395

AAT GCT CCT CCG AAT CCG ATG GCT TTT GGT GTT CCA AAC ATG CAT AAC      2031
Asn Ala Pro Pro Asn Pro Met Ala Phe Gly Val Pro Asn Met His Asn
            400                 405                 410

CTT TCT GGA CCA GCA GTA TCT CAA CCG TTT TCC TTG CCT CCT GCT CCT      2079
Leu Ser Gly Pro Ala Val Ser Gln Pro Phe Ser Leu Pro Pro Ala Pro
        415                 420                 425

TTA CCG AGG GAC TCT GGT TAC AGC AGC TCC TCC CCT GGG CAG TTG TTA      2127
Leu Pro Arg Asp Ser Gly Tyr Ser Ser Ser Ser Pro Gly Gln Leu Leu
    430                 435                 440

GAT ATA CTA AAT TCG AAA AAG CCT GAC AGC AAC GTG CAA TCA AGC AAA      2175
Asp Ile Leu Asn Ser Lys Lys Pro Asp Ser Asn Val Gln Ser Ser Lys
445                 450                 455                 460

AAG CCA AAG CTT AAA ATC TTA CAG AGA GGA ACG GAC TTG AAT TCA CTC      2223
Lys Pro Lys Leu Lys Ile Leu Gln Arg Gly Thr Asp Leu Asn Ser Leu
                465                 470                 475

AAG CAA AAC AAT AAT GAT GAA ACT GCT CAT TCA AAC TCT CAA GCT TTG      2271
Lys Gln Asn Asn Asn Asp Glu Thr Ala His Ser Asn Ser Gln Ala Leu
            480                 485                 490

CTA GAT TTG TTG AAA AAA CCA ACA TCA TCG CAG AAG ATA CAC GCT TCC      2319
Leu Asp Leu Leu Lys Lys Pro Thr Ser Ser Gln Lys Ile His Ala Ser
        495                 500                 505

AAA CCA GAT ACT TCC TTT TTA CCA AAT GAC TCC GTA TCT GGT ATA CAA      2367
Lys Pro Asp Thr Ser Phe Leu Pro Asn Asp Ser Val Ser Gly Ile Gln
    510                 515                 520

GAT GCA GAA TAT GAA GAT TTC GAG AGT AGT TCA GAT GAA GAG GTG GAG      2415
Asp Ala Glu Tyr Glu Asp Phe Glu Ser Ser Ser Asp Glu Glu Val Glu
525                 530                 535                 540

ACA GCT AGA GAT GAA AGA AAT TCA TTG AAT GTA GAT ATT GGG GTG AAC      2463
Thr Ala Arg Asp Glu Arg Asn Ser Leu Asn Val Asp Ile Gly Val Asn
                545                 550                 555

GTT ATG CCA AGC GAA AAA GAC AGC CGA AGA AGT CAA AAG GAA AAA CCA      2511
Val Met Pro Ser Glu Lys Asp Ser Arg Arg Ser Gln Lys Glu Lys Pro
            560                 565                 570

AGG AAC GAC GCA AGC AAA ACA AAC TTG AAC GCT TCT GCA GAA TCT AAT      2559
Arg Asn Asp Ala Ser Lys Thr Asn Leu Asn Ala Ser Ala Glu Ser Asn
        575                 580                 585

AGT GTA GAA TGG GGG GCT GGG TAAATCTTCA CCCTCCGACT TCAGAGTAAC        2610
Ser Val Glu Trp Gly Ala Gly
    590                 595

ACAGAATCCA CAGTA                                                    2625
```

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6854 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2050..4053

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AGCTTCTCCC TTTTCCTTCA GTGCTGCTAC TCTCTGCTCT CCACTTAAGT GTTACAATTA 60

ATTTGCAGCT AGTTTGCAGT TCGTACAACC TCGCCTATTC TTGTAACGAA GAAGAACGTA 120

TTTATAATAT TGGGCTGTAA TGTGTTGAGT TTAGTAATAG ATAAAGTAGG ACAGAGTTCT 180

GTCTTTGTTT ATCTATGGGG TTCAGAGTGA TAAGGGGCAG GATAAGGAAG TTAAAAAAAA 240

AAAGGTTACG TTATATAACG AAAGAAAAGA AACGAGCGAA GTGCCAACTA TAGCCCAATA 300

TCAAGAATGC AAGTCAGCAA AGTACAGTAA TCGTATGAAG ATACGCGATG CGTAATATCC 360

CTCAAGGGCT CCGGATCAGA AAAGCTAAGG GAAGATCCTT ACATTACACG GCGTGCGACA 420

GACTCGAACC ACAGCTAACT TCTCGTGAAA AGATGGCTTC AACTTCGCTC TTGCAATAAC 480

TTTGAAACAC ACGAACAAAG GTTTATTGCG CTTGATTAAC GTTGGAAGTA TATGATACTA 540

ATACTACTTT GTTCTCTAAG TCATCGCTAT ATGTTTATCT CGAGGAAAAG GTGCACGGCG 600

GTACACAATT ACTTCGCCGT TTCGGGTAAA ACAAGTGTTA CATTTATAAT ATATATGTAT 660

ATATGTATGT GCGCGTAAGT ATATGCCGTT CATAACAAAT CATCTTCTTG TTGCTGGATG 720

GACTCCTTAA TTTTATTCAA AATGGTAATT TTCCATTTAT CTAGTCTCAT AAAATTGTCA 780

AACTCCTTAC AGTGTTCGCT TAGCTGCTCG CTATCACCTT CATTAACAGC ATCGATTAAA 840

CTTTTCAAGA AATTTGACTC CCTTGAATCC GCAAAATTCG GATCTTCACT TTGACCCTCT 900

TGTAAAGTTC TTGCAGCAGC GACTGCATCA GTAGCAGCTA GCTGACAAAG CCCTTTTTTT 960

AGGAAGTAAT CCTTCAAACT CCATTGGCTC AATCTATTGC CCATGCTGCT CTTGATCAAC 1020

TTCGAATATA TATCACTTGC TTCAATATAT TGACCGTCAA GAGCCTTTAG ATCTGCGCAT 1080

TTGATAAAAC ACTTATTCGA TAATGCTACC GACTGGTCTT GGGCATACCA CTCACCAGCG 1140

AGCTCATAGC AATCTATAGC TTTTGCATAG TCATGCAAAT CATTTTCTAG AATTTCTCCA 1200

AGCTCAAACT TGAAATTAGC ACCTCTCCGG AACTGCCCCC TATGAGTAAA AATTTGAATA 1260

GCATTTTCTA ATGAATCCAC GGCGTTCACA GAGTTTCCAC CGCTTTTAAA GCATTTATAA 1320

GCCTCTACGT AGGTATTTCC TGCTTCGTCT TCATTACCAG CCTTTTTCTG ATAGTCAGCA 1380

```
GCTTTCAAAA ACGAGTCTCC TGCCAAGTTT AACTCTTTTC TTAGACGGTA AATGGTGGCT    1440

GCTTGGACAC AAAGATCAGC AGCCTCCTCA AACTTGTATG AATCAGAACC GCTAAACAAT    1500

TTCATGAAAC CCGATGAAGG AACACCCTTC TTCTCAGCCT TAACACAACG GGAAATATCA    1560

ATTCCCGTAT TTCAATGTTA GTAATTTGCC TTCGTAAATT ACGGAATCAC ATAGCTTTCA    1620

TTTTGTTCCT TTGATATATT TCCCTACTAC ATACTCTTTT CAATAACTCT ACAGGGTCTG    1680

ACATTTTTAA CTTTCAGGTT AATGATGGTG TTCTTACTAT ATTCTCGAGT CGTACAGAAG    1740

TTAGTTCAGA TAAACTGCTT CGGTGCTGCC CACTTCTTAT CATTACTTCA ACTTTACCTT    1800

CCCTATACCT GTGTGTCCTT ATTAATTCAA GTTAATCCGA GGTAATAGAT TAGGGTAACC    1860

TTCAATGATG TCACGAAACA CGGATGCTGC AACTTTGCGA TTTTTTCCTG GAAAAGAATA    1920

ACAATTAAAG GCAGCCTTTC AGCTGAGATT ACCAGCAGGT CTTTGGAGAT TAGCGCAAGA    1980

AGAAGTGTGA TATAGTACTC ATAGAGGCAG GCTACAGACT AGGGAAAGCG TGTTCAACAA    2040

CAATAAGAA ATG GAG ACC AGT TCT TTT GAG AAT GCT CCT CCT GCA GCC          2088
          Met Glu Thr Ser Ser Phe Glu Asn Ala Pro Pro Ala Ala
            1               5                  10

ATC AAT GAT GCT CAG GAT AAT AAT ATA AAT ACG GAG ACT AAT GAC CAG      2136
Ile Asn Asp Ala Gln Asp Asn Asn Ile Asn Thr Glu Thr Asn Asp Gln
     15                  20                  25

GAA ACA AAT CAG CAA TCT ATC GAA ACT AGA GAT GCA ATT GAC AAA GAA      2184
Glu Thr Asn Gln Gln Ser Ile Glu Thr Arg Asp Ala Ile Asp Lys Glu
 30                  35                  40                  45

AAC GGT GTG CAA ACG GAA ACT GGT GAG AAC TCT GCA AAA AAT GCC GAA      2232
Asn Gly Val Gln Thr Glu Thr Gly Glu Asn Ser Ala Lys Asn Ala Glu
                 50                  55                  60

CAA AAC GTT TCT TCT ACA AAT TTG AAT AAT GCC CCC ACC AAT GGT GCT      2280
Gln Asn Val Ser Ser Thr Asn Leu Asn Asn Ala Pro Thr Asn Gly Ala
            65                  70                  75

TTG GAC GAT GAT GTT ATC CCA AAT GCT ATT GTT ATT AAA AAC ATT CCG      2328
Leu Asp Asp Asp Val Ile Pro Asn Ala Ile Val Ile Lys Asn Ile Pro
         80                  85                  90

TTT GCT ATT AAA AAA GAG CAA TTG TTA GAC ATT ATT GAA GAA ATG GAT      2376
Phe Ala Ile Lys Lys Glu Gln Leu Leu Asp Ile Ile Glu Glu Met Asp
     95                  100                 105

CTT CCC CTT CCT TAT GCC TTC AAT TAC CAC TTT GAT AAC GGT ATT TTC      2424
Leu Pro Leu Pro Tyr Ala Phe Asn Tyr His Phe Asp Asn Gly Ile Phe
110                 115                 120                 125

AGA GGA CTA GCC TTT GCG AAT TTC ACC ACT CCT GAA GAA ACT ACT CAA      2472
Arg Gly Leu Ala Phe Ala Asn Phe Thr Thr Pro Glu Glu Thr Thr Gln
                130                 135                 140

GTG ATA ACT TCT TTG AAT GGA AAG GAA ATC AGC GGG AGG AAA TTG AAA      2520
Val Ile Thr Ser Leu Asn Gly Lys Glu Ile Ser Gly Arg Lys Leu Lys
            145                 150                 155
```

```
GTG GAA TAT AAA AAA ATG CTT CCC CAA GCT GAA AGA GAA AGA ATC GAG      2568
Val Glu Tyr Lys Lys Met Leu Pro Gln Ala Glu Arg Glu Arg Ile Glu
        160                 165                 170

AGG GAG AAG AGA GAG AAA AGA GGA CAA TTA GAA GAA CAA CAC AGA TCG      2616
Arg Glu Lys Arg Glu Lys Arg Gly Gln Leu Glu Glu Gln His Arg Ser
    175                 180                 185

TCA TCT AAT CTT TCT TTG GAT TCT TTA TCT AAA ATG AGT GGA AGC GGA      2664
Ser Ser Asn Leu Ser Leu Asp Ser Leu Ser Lys Met Ser Gly Ser Gly
190                 195                 200                 205

AAC AAT AAT ACT TCT AAC AAT CAA TTA TTC TCG ACT CTA ATG AAC GGC      2712
Asn Asn Asn Thr Ser Asn Asn Gln Leu Phe Ser Thr Leu Met Asn Gly
                210                 215                 220

ATT AAT GCT AAT AGC ATG ATG AAC AGT CCA ATG AAT AAT ACC ATT AAC      2760
Ile Asn Ala Asn Ser Met Met Asn Ser Pro Met Asn Asn Thr Ile Asn
            225                 230                 235

AAT AAC AGT TCT AAT AAC AAC AAT AGT GGT AAC ATC ATT CTG AAC CAA      2808
Asn Asn Ser Ser Asn Asn Asn Asn Ser Gly Asn Ile Ile Leu Asn Gln
        240                 245                 250

CCT TCA CTT TCT GCC CAA CAT ACT TCT TCA TCG TTG TAC CAA ACA AAC      2856
Pro Ser Leu Ser Ala Gln His Thr Ser Ser Ser Leu Tyr Gln Thr Asn
    255                 260                 265

GTT AAT AAT CAA GCC CAG ATG TCC ACT GAG AGA TTT TAT GCG CCT TTA      2904
Val Asn Asn Gln Ala Gln Met Ser Thr Glu Arg Phe Tyr Ala Pro Leu
270                 275                 280                 285

CCA TCA ACT TCC ACT TTG CCT CTC CCA CCC CAA CAA CTG GAC TTC AAT      2952
Pro Ser Thr Ser Thr Leu Pro Leu Pro Pro Gln Gln Leu Asp Phe Asn
                290                 295                 300

GAC CCT GAC ACT TTG GAA ATT TAT TCC CAA TTA TTG TTA TTT AAG GAT      3000
Asp Pro Asp Thr Leu Glu Ile Tyr Ser Gln Leu Leu Leu Phe Lys Asp
            305                 310                 315

AGA GAA AAG TAT TAT TAC GAG TTG GCT TAT CCC ATG GGT ATA TCC GCT      3048
Arg Glu Lys Tyr Tyr Tyr Glu Leu Ala Tyr Pro Met Gly Ile Ser Ala
        320                 325                 330

TCC CAC AAG AGA ATT ATC AAT GTT TTG TGC TCG TAC TTA GGG CTA GTA      3096
Ser His Lys Arg Ile Ile Asn Val Leu Cys Ser Tyr Leu Gly Leu Val
    335                 340                 345

GAA GTA TAT GAT CCA AGA TTT ATT ATT ATC AGA AGA AAG ATT CTG GAT      3144
Glu Val Tyr Asp Pro Arg Phe Ile Ile Ile Arg Arg Lys Ile Leu Asp
350                 355                 360                 365

CAT GCT AAT TTA CAA TCT CAT TTG CAA CAA CAA GGT CAA ATG ACA TCT      3192
His Ala Asn Leu Gln Ser His Leu Gln Gln Gln Gly Gln Met Thr Ser
                370                 375                 380

GCT CAT CCT TTG CAG CCA AAC TCC ACT GGC GGC TCC ATG AAT AGG TCA      3240
Ala His Pro Leu Gln Pro Asn Ser Thr Gly Gly Ser Met Asn Arg Ser
            385                 390                 395

CAA TCT TAT ACA AGT TTG TTA CAG GCC CAT GCA GCA GCT GCA GCG AAT      3288
Gln Ser Tyr Thr Ser Leu Leu Gln Ala His Ala Ala Ala Ala Ala Asn
        400                 405                 410
```

```
AGT ATT AGC AAT CAG GCC GTT AAC AAT TCT TCC AAC AGC AAT ACT ATT      3336
Ser Ile Ser Asn Gln Ala Val Asn Asn Ser Ser Asn Ser Asn Thr Ile
    415                 420                 425

AAC AGT AAT AAC GGT AAC GGT AAC AAT GTC ATC ATT AAT AAC AAT AGC      3384
Asn Ser Asn Asn Gly Asn Gly Asn Asn Val Ile Ile Asn Asn Asn Ser
430                 435                 440                 445

GCC AGC TCA ACA CCA AAA ATT TCT TCA CAG GGA CAA TTC TCC ATG CAA      3432
Ala Ser Ser Thr Pro Lys Ile Ser Ser Gln Gly Gln Phe Ser Met Gln
                450                 455                 460

CCA ACA CTA ACC TCA CCT AAA ATG AAC ATA CAC CAT AGT TCT CAA TAC      3480
Pro Thr Leu Thr Ser Pro Lys Met Asn Ile His His Ser Ser Gln Tyr
            465                 470                 475

AAT TCC GCA GAC CAA CCG CAA CAA CCT CAA CCA CAA ACA CAG CAA AAT      3528
Asn Ser Ala Asp Gln Pro Gln Gln Pro Gln Pro Gln Thr Gln Gln Asn
        480                 485                 490

GTT CAG TCA GCT GCG CAA CAA CAA CAA TCT TTT TTA AGA CAA CAA GCT      3576
Val Gln Ser Ala Ala Gln Gln Gln Gln Ser Phe Leu Arg Gln Gln Ala
    495                 500                 505

ACT TTA ACA CCA TCC TCA AGA ATT CCA TCC GGT TAT TCT GCC AAC CAT      3624
Thr Leu Thr Pro Ser Ser Arg Ile Pro Ser Gly Tyr Ser Ala Asn His
510                 515                 520                 525

TAT CAA ATC AAT TCC GTT AAT CCC TTA CTG AGA AAT TCT CAA ATT TCA      3672
Tyr Gln Ile Asn Ser Val Asn Pro Leu Leu Arg Asn Ser Gln Ile Ser
                530                 535                 540

CCT CCA AAT TCA CAA ATC CCA ATC AAC AGC CAA ACC CTA TCC CAA GCG      3720
Pro Pro Asn Ser Gln Ile Pro Ile Asn Ser Gln Thr Leu Ser Gln Ala
            545                 550                 555

CAA CCA CCA GCA CAG TCC CAA ACT CAA CAA CGG GTA CCA GTG GCA TAC      3768
Gln Pro Pro Ala Gln Ser Gln Thr Gln Gln Arg Val Pro Val Ala Tyr
        560                 565                 570

CAA AAT GCT TCA TTG TCT TCC CAG CAG TTG TAC AAC CTT AAC GGC CCA      3816
Gln Asn Ala Ser Leu Ser Ser Gln Gln Leu Tyr Asn Leu Asn Gly Pro
    575                 580                 585

TCT TCA GCA AAC TCA CAG TCC CAA CTG CTT CCA CAG CAC ACA AAT GGC      3864
Ser Ser Ala Asn Ser Gln Ser Gln Leu Leu Pro Gln His Thr Asn Gly
590                 595                 600                 605

TCA GTA CAT TCT AAT TTC TCA TAT CAG TCT TAT CAC GAT GAG TCC ATG      3912
Ser Val His Ser Asn Phe Ser Tyr Gln Ser Tyr His Asp Glu Ser Met
                610                 615                 620

TTG TCC GCA CAC AAT TTG AAT AGT GCC GAC TTG ATC TAT AAA TCT TTG      3960
Leu Ser Ala His Asn Leu Asn Ser Ala Asp Leu Ile Tyr Lys Ser Leu
            625                 630                 635

AGT CAC TCT GGA CTA GAT GAT GGC TTG GAA CAG GGC TTG AAT CGT TCT      4008
Ser His Ser Gly Leu Asp Asp Gly Leu Glu Gln Gly Leu Asn Arg Ser
        640                 645                 650

TTA AGC GGA CTG GAT TTA CAA AAC CAA AAC AAG AAG AAT CTA TGG          4053
Leu Ser Gly Leu Asp Leu Gln Asn Gln Asn Lys Lys Asn Leu Trp
    655                 660                 665
```

TAATATATAC TTCCATTATT CTATGATTAT AGAGTTTGTT TGGTATTTGT ATATCGCACG 4113

ATACAAGTAA TGAGGGGTGC TTACACAAGA TAAAAGATAA AAAAATATAT ATATATAATA 4173

AAAACCATCA AAAACACCAT TGAAAAAAAA TATAAAAAAA AAAAAAAATA ACCGAATATG 4233

AATATGAAAT TAATGATCAT GATGAAGTTA ATTTTTACTG AGAAACGTCA CCTAATGTCG 4293

ATGAAACGAT GATAATGAAT GAATGATGAG GCTACTTTAA GTAACGCAAT GTAATCAAGC 4353

CAAAATTATC CCTCTTTTTT TTTTTTCCCT CTTTTGAGAT TTTATTTTTA ACCTACTACT 4413

TACTTTTTTT TTTTGAACGT TCTTTTCCCA CATACTTTTA TATATGGTAT TTATATGTAC 4473

GATGTTTAAT CACAGAGATG TTTCTACCTT ACTCGATATT GTTTTTGCAT TAATTGATAT 4533

CTTGCTCACT GCATCATTGG CGGTATTTGT AGTATATAGA AAGTCGGGTA ACAATAATTT 4593

ATTGACATTT CTTTGTTTAC AATGATCAGA GAAGAGCAGA AAGTTTCATA GTCAAACGTT 4653

CAGGCCAATT GAACAAGAAA TTATTCGTTT TTTTAGTCGT TGAGTGTTCA ACTGACATGC 4713

TATTTTGGTG GTTCTTGATT AATTGGGGGC TTCATTGTTT GAAATAAAGA GTCGGGAAAA 4773

TAGCACAGAA ACAAAGCATA TTAAAAGAGG CAAAAGAAGA AAGAACGAAT ATAAAAGGTA 4833

AAAAAGGAAA AGCATTGCTA TTCTTTTCTC ATAGGTGTTA TTCATACCGC CCTCTCTCTT 4893

CTTCCTTCTT CATTAATTAG TCTCCGTATA ATTTGCAGAT AATGTCATTA ACAGCAAACG 4953

ACGAATCGCC AAAACCCAAA AAAAATGCAT TATTGAAAAA CTTAGAGATC GATGATCTGA 5013

TACATTCTCA ATTTGTCAGA AGCGATACAA ATGGACATAG AACTACAAGA CGACTATTCA 5073

ACTCCGATGC CAGTATATCA CATCGAATAA GAGGAAGTGT TCGGTCTGAT AAAGGCCTTA 5133

ATAAAATAAA AAAAGGGTTG ATTTCCCAGC AGTCCAAACT TGCGTCAGAA AATTCTTCTC 5193

AAAATATCGT TAATAGGGAC AATAAGATGG GAGCAGTAAG TTTCCCCATT ATTGAACCTA 5253

ATATTGAAGT CAGCGAGGAG TTGAAGGTTA GAATTAAGTA TGATTCTATC AAATTTTTCA 5313

ATTTTGAAAG ACTAATATCT AAATCTTCAG TCATAGCACC TTTAGTTAAC AAAAATATAA 5373

CATCATCCGG TCCTCTAATC GGGTTTCAAA GAAGAGTTAA CAGGTTAAAG CAAACATGGG 5433

ATCTAGCAAC CGAAAACATG GAGTACCCAT ATTCTTCTGA TAATACGCCA TTCAGGGATA 5493

ACGATTCTTG GCAATGGTAC GTACCATACG GCGGAACAAT AAAAAAAATG AAAGATTTCA 5553

GTACAAAAAG AACTTTACCC ACCTGGGAAG ATAAAATAAA GTTTCTTACA TTTTTAGAAA 5613

ACTCTAAGTC TGCAACGTAC ATTAATGGTA ACGTATCACT TTGCAATCAT AATGAAACCG 5673

ATCAAGAAAA CGAAGATAGG AAAAAAAGGA AAGGGAAAGT ACCAAGAATC AAAAATAAAG 5733

TGTGGTTTTC CCAGATAGAA TACATTGTTC TTCGAAATTA TGAAATTAAA CCTTGGTATA 5793

CATCTCCTTT TCCGGAACAC ATCAACCAAA ATAAAATGGT TTTTATATGT GAGTTCTGCC 5853

TAAAATATAT GACTTCTCGA TATACTTTTT ATAGACACCA ACTAAAGTGT CTAACTTTTA 5913

AGCCCCCCGG AAATGAAATT TATCGCGACG GTAAGCTGTC TGTTTGGGAA ATTGATGGGC 5973

GGGAGAATGT CTTGTATTGT CAAAATCTTT GCCTGTTGGC AAAATGTTTT ATCAATTCTA 6033

AGACTTTGTA TTACGATGTT GAACCGTTTA TATTCTATAT TCTAACGGAG AGAGAGGATA 6093

CAGAGAACCA TCCCTATCAA AACGCAGCCA AATTCCATTT CGTAGGCTAT TTCTCCAAGG 6153

AAAAATTCAA CTCCAATGAC TATAACCTAA GTTGTATTTT AACTCTACCC ATATACCAGA 6213

GGAAAGGATA TGGTCAGTTT TTGATGGAAT TTTCATATTT ATTATCCAGA AAGGAGTCAA 6273

AATTTGGAAC TCCTGAAAAA CCATTGTCGG ATTTAGGATT ATTGACTTAC AGAACGTTTT 6333

GGAAGATAAA ATGTGCTGAA GTGCTATTAA AATTAAGAGA CAGTGCTAGA CGTCGATCAA 6393

ATAATAAAAA TGAAGATACT TTTCAGCAGG TTAGCCTAAA CGATATCGCT AAACTAACAG 6453

GAATGATACC AACAGACGTT GTGTTTGGAT TGGAACAACT TCAAGTTTTG TATCGCCATA 6513

AAACACGCTC ATTATCCAGT TTGGATGATT TCAACTATAT TATTAAAATC GATTCTTGGA 6573

ACAGGATTGA AAATATTTAC AAAACTTGGA GCTCAAAAAA CTATCCTCGC GTCAAATATG 6633

ACAAACTATT GTGGGAACCT ATTATATTAG GGCCGTCATT TGGTATAAAT GGGATGATGA 6693

ACTTAGAACC CACCGCATTA GCGGACGAAG CTCTTACAAA TGAAACTATG GCTCCGGTAA 6753

TTTCGAATAA CACACATATA GAAAACTATA ACAACAGTAG AGCACATAAT AAACGCAGAA 6813

GAAGAAGAAG AAGAAGTAGT GAGCACAAAA CATCCAAGCT T 6854

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2814 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..696

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
GAA TTC CAA TAC ACC AAA CAG CTG CAT TTC CCT GTG GGG CCC AAA TCC        48
Glu Phe Gln Tyr Thr Lys Gln Leu His Phe Pro Val Gly Pro Lys Ser
  1               5                  10                  15

ACA AAC TGT GAG GTA GCG GAA ATT CTT TTA CAC TGC GAC TGG GAA AGG        96
Thr Asn Cys Glu Val Ala Glu Ile Leu Leu His Cys Asp Trp Glu Arg
            20                  25                  30

TAC ATA AAT GTT TTA AGT ATA ACA AGA ACA CCA AAT GTT CCT AGT GGT       144
Tyr Ile Asn Val Leu Ser Ile Thr Arg Thr Pro Asn Val Pro Ser Gly
        35                  40                  45

ACC AGT TTC AGC ACC AGA ACG AGG TAC ATG TTC CGA TGG GAT GAC CAG       192
Thr Ser Phe Ser Thr Arg Thr Arg Tyr Met Phe Arg Trp Asp Asp Gln
    50                  55                  60
```

```
GGG CAA GGT TGC ATA TTA AAA ATA AGT TTT TGG GTG GAC TGG AAC GCA       240
Gly Gln Gly Cys Ile Leu Lys Ile Ser Phe Trp Val Asp Trp Asn Ala
 65                  70                  75                  80

TCC AGT TGG ATC AAG CCA ATG GTA GAG AGC AAT TGT AAA AAT GGA CAA       288
Ser Ser Trp Ile Lys Pro Met Val Glu Ser Asn Cys Lys Asn Gly Gln
                 85                  90                  95

ATT AGC GCC ACT AAG GAC TTG GTA AAG TTA GTC GAA GAA TTT GTA GAG       336
Ile Ser Ala Thr Lys Asp Leu Val Lys Leu Val Glu Glu Phe Val Glu
            100                 105                 110

AAA TAC GTG GAA TTG AGC AAA GAA AAA GCA GAT ACA CTC AAG CCG TTG       384
Lys Tyr Val Glu Leu Ser Lys Glu Lys Ala Asp Thr Leu Lys Pro Leu
        115                 120                 125

CCC AGT GTT ACA TCT TTT GGA TCA CCT AGG AAA GTG GCA GCA CCG GAG       432
Pro Ser Val Thr Ser Phe Gly Ser Pro Arg Lys Val Ala Ala Pro Glu
    130                 135                 140

CTG TCG ATG GTA CAG CCG GAG TCG AAA CCA GAA GCT GAG GCG GAA ATC       480
Leu Ser Met Val Gln Pro Glu Ser Lys Pro Glu Ala Glu Ala Glu Ile
145                 150                 155                 160

TCA GAA ATA GGC AGC GAC AGA TGG AGG TTT AAC TGG GTG AAC ATA ATA       528
Ser Glu Ile Gly Ser Asp Arg Trp Arg Phe Asn Trp Val Asn Ile Ile
                165                 170                 175

ATC TTG GTG CTC TTG GTG TTA AAT CTG CTG TAT TTA ATG AAG TTG AAC       576
Ile Leu Val Leu Leu Val Leu Asn Leu Leu Tyr Leu Met Lys Leu Asn
            180                 185                 190

AAG AAG ATG GAT AAG CTG ACG AAC CTC ATG ACC CAC AAG GAC GAA GTT       624
Lys Lys Met Asp Lys Leu Thr Asn Leu Met Thr His Lys Asp Glu Val
        195                 200                 205

GTA GCG CAC GCG ACT CTA TTG GAC ATA CCA GCC CAA GTA CAA TGG TCA       672
Val Ala His Ala Thr Leu Leu Asp Ile Pro Ala Gln Val Gln Trp Ser
    210                 215                 220

AGA CCA AGA AGG GGA GAC GTG TTG TAACAGAGTA ATCATGTAAT ATTGTATGTA      726
Arg Pro Arg Arg Gly Asp Val Leu
225                 230

AGGTTATGTA TGTTCGTATG GTATGGAAAA AAAAAAAAAA AAAGGATGCT ATGTGGAGAA     786

TGTAAGGCGT GGTAGCTCCG GATAATTCAG TCTGTAGGCT TCATCACGGG CAGTGGCCTG     846

ACTCTGAGAG CTTGCTCCGG TATTAAGTTG TGCGTTTGAA ATTTTCTGGA AAAAAGAAAT     906

TGATTGGTTG AAGCTATACT CGTCGAAAGA TTTCTTCGGC AGTGGTTGTT GCTCCACCTG     966

CACGGGAGTT GTGTTTGCGT TTATGTTCGG CTTGGCTATA TTATTAGCGA GTGATGTTTG    1026

CAATTTGCTG TATTGAGAAT CAATTTGGGT GCGTAAGCTT TCAATAATTT TGCAGACCGC    1086

AGGCACTTCC AACTTTATGA GTTGCAGGTA TTCTCTTTTA TGAATATACG ATGACGACGA    1146

TGACGACGAC GCATCCATGC GCAAAAGCTC AGGGTGTCTA GATAGTTTGT TAGTCAATAA    1206

ATCCACATAT CTAAAATAAT AAATAAACGA CAGCGACAAG TCGTTGGCCT GGAACGCACA    1266

CTGTGCCTTT TCCAATATGC CGATGCATGT TTTCAGGTAA ATTCTCAATG GTATCGCCGG    1326
```

ATTGAAGCGA TAATCCTTAG CGTCCTGAAC CAATTGCTTA CTAGACTTCA TGACCTACCG 1386

GGGCCAGATA AAGATGCGGA AGGAAGAGAA AAAATGTATA GTGGTTGGTG AACCGCAACA 1446

ATAATTCGTG CCAACACTTT AATCGAAGCA AAAATTGTCT TGTATGTTAT TAATATTATC 1506

TATCTAACCA TTGATTTACG TATAAAACTG TCGATGCTCA TCGCCTAGCA ATGAAAAAAT 1566

TTTTTCTTTT TTTTTTCATT ATTTCTCTTT GTTGCGTACT TTTTTTCATT GCGTTTCGCG 1626

GCAAAAGCGA TTCGAGTTGA CTGGAAGTGT GTTATACTAT AAAAAGTGTA TATGCCTATT 1686

TTTGGTTCTG ATCTTTACTT TACTGTTAAG TACTGGCTGA GGCAGTAGAC TCTGCCTCTG 1746

TTACGGCAGC GGTATTCGCC TCGGCATCAG CAGCCGCCCA CGGTAGAGTA GGTTCTGTTG 1806

TTTTGACGTT TGCCAAGGTA CTGTCCAAAT GCTCCTTCAG CAAGGCCTCA TTACTTTCCT 1866

TCTCCGGACC CACCGATTGC GTGATCTCCT GTACACGGTT CAAGAACTTG TTCAAATTGT 1926

AGCCCGCAGC AGCATCAGAG ACTTCTTGTG TGTAAGGGAC ACCCCTCAAC TCCTTGACTC 1986

TTCTTTTGTG CACTTTGCCC TTTAAATGCG TTTTTAACGC TATAGCAGTC TCCATGTATT 2046

TGGCACAGTG TATGCAATAG TGCTGACCAA GGCCCGGTTT GGTTTCATCC AATGGCTGGT 2106

TCAGAAGCTT CTGTACTGAT TCCTTGGTGG ACAAATCGTT ATAGATCAGG TCCAAGTCTC 2166

GTGTTCTTCT TTTAGTCTTG TATCTCTTCA CCGAATATCT ACCCATGATG CGCTATTGTT 2226

TTATCTTCAC TTGTCTGTGT GTTTAACTGC CTTTCAATTC ACCTCATCTC ATCTCCCGCT 2286

ACTTTCCATA TATAAAAGCA AAATTAATTT GCTTTTTCCC CTGTCAGTAT AAAAAAATTT 2346

TCCGCAGGAT ATAGAAAAAA AAGAAATGAA ATTATAGTAG CGGTTATTTC CGTGGGGTGC 2406

TTTTTTACAC CTGTACATCT TTTCCCTCCG TACATTTTTT TTATTTTTTT TTTGGGTTTT 2466

TTTTTTTCGA TATTTTTCCC TCCGAAACTA GTTAGCACAA TAATGCTGAC TAAGGAAACT 2526

TTTCATCTCA GAATTGATGG TCAGTTTGGT TTCTCTAGAG AATAGTTTAT AAAAAGATGT 2586

TGATGTGGAG CAACCATTTA TACATCCTTT CCGCAAGTGC TTTTGGAGTG GGACTTTCAA 2646

ACTTTAAAGT ACAGTATATC AAATAACTAA TTCAAGATGG CTAGAAGACC AGCTAGATGT 2706

TACAGATACC AAAAGAACAA GCCTTACCCA AAGTCTAGAT ACAACAGAGC TGTTCCAGAC 2766

TCCAAGATCA GAATCTACGA TTTGGGTAAG AAGAAGGCTA CCGTCGAT 2814

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ile Asn Leu Lys Ala Leu Ala Ala Leu Ala Lys Lys Ile Leu
1 5 10

WHAT IS CLAIMED IS:

1. A host cell transformed or transfected with DNA comprising:

a repressor gene encoding a repressor protein, said repressor gene under transcriptional control of a promoter;

a selectable marker gene encoding a selectable marker protein; said selectable marker gene under transcriptional control of an operator; said operator regulated by interaction with said repressor protein;

a first recombinant fusion protein gene encoding a first binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activating protein or a transactivating domain of a transcriptional activating protein; and

a second recombinant fusion protein gene encoding a second binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activating protein or a transactivating domain of a transcriptional activating protein, whichever domain is not encoded by the first fusion protein gene, said second binding protein or binding fragment thereof capable of interacting with said first binding protein or binding fragment thereof such that interaction of said second binding protein or binding fragment thereof and said first binding protein or binding fragment thereof brings into proximity a DNA binding domain and a transactivating domain forming a functional transcriptional activating protein; said functional transcriptional activating protein acting on said promoter to increase expression of said repressor gene.

2. The host cell of claims 1 wherein said DNA binding domain and said transactivating domain are derived from a common transcriptional activating protein.

3. The host cell of claim 1 wherein one or more of the repressor gene, the selectable marker gene, the first recombinant fusion protein gene, and the second recombinant fusion protein gene are encoded on distinct DNA expression constructs.

4. The host cell of claim 1 wherein said selectable marker protein is an enzyme in a pathway for synthesis of a nutritional requirement for said host cell such that expression of said selectable marker protein is required for growth of said host cell on media lacking said nutritional requirement.

5. The host cell of claim 1 wherein said host cell is a yeast cell or a mammalian.

6. The host cell of claim 2 wherein said selectable marker gene encodes HIS3;

7. The host cell of claim 2 wherein said repressor protein gene encodes a tetracycline resistance protein;

8. The host cell of claim 2 wherein said operator is a *tet* operator.

9. The host cell of claim 2 wherein said promoter is selected from the group consisting of the LexA promoter, the alcohol dehydrogenase promoter, the Gal4 promoter.

10. The host cell of claim 2 wherein said DNA binding domain derived from a protein selected from the group consisting of LexA and Gal4.

11. The host cells of claim 2 wherein said transactivating domain is derived from a protein selected from the group consisting of VP16 and Gal4.

12. The host cell of claim 2 wherein the first binding protein is CREB and the second binding protein is CBP.

13. The host cell of claim 2 wherein the first binding protein is Tax and the second binding protein is SRF.

14. The host cell of claim 2 wherein the first binding protein is casein kinase I and the second binding protein is CREB.

15. The host cell of claim 2 wherein the first binding protein is AKAP 79 and the second binding protein is selected from the group consisting of RI, RII and calcineurin.

16. A method to identify an inhibitor of binding between a first binding protein or binding fragment thereof and a second binding protein or binding fragment thereof comprising the steps of:

a) growing host cells of any one of claims 1 through 15 in the absence of a test compound and under conditions which permit expression of said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof such that said first binding protein or fragment thereof and second binding protein or binding fragment thereof interact bringing

into proximity said DNA binding domain and said transactivating domain forming said functional transcriptional activating protein; said transcriptional activating protein acting on said promoter to increase expression of said repressor protein; said repressor protein interacting with said operator such that said selectable marker protein is not expressed;

b) confirming lack of expression of said selectable marker protein in said host cell;

c) growing said host cells in the presence of a test compound; and

d) comparing expression of said selectable marker protein in the presence and absence of said test compound wherein increased expression of said selectable marker protein is indicative that the test compound is an inhibitor of binding between said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof.

17. The method of claim 16 wherein

the host cell is a yeast cell;

the selectable marker gene encodes HIS3;

transcription of the selectable marker gene is regulated by the *tet* operator;

the repressor protein gene encodes the tetracycline resistance protein;

transcription of the tetracycline resistance protein is regulated by the LexA promoter;

the DNA binding domain is derived from LexA; and

the transactivating domain is derived from VP16.

18. The method of claim 16 wherein

the host cell is a yeast cell;

the selectable marker gene encodes HIS3;

transcription of the selectable marker gene is regulated by the *tet* operator;

the repressor protein gene encodes the tetracycline resistance protein;

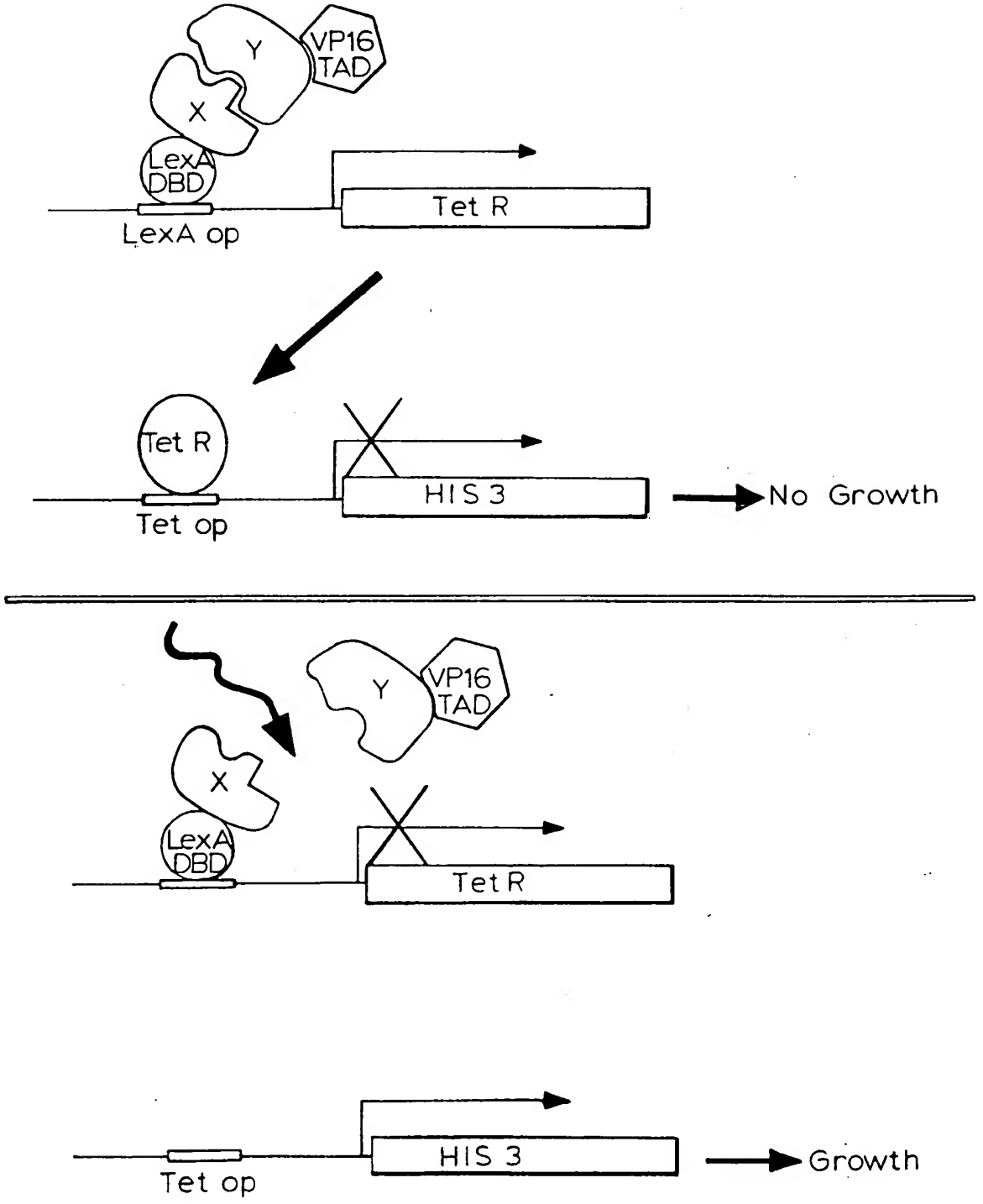
transcription of the tetracycline resistance protein is regulated by the alcohol dehydrogenase promoter;

the DNA binding domain is derived from LexA; and

the transactivating domain is derived from VP16.

19. A kit to identify an inhibitor of binding between a first binding protein or binding fragment thereof and a second binding protein or binding fragment thereof, said inhibitor identified by the method of claim 16.

FIGURE 1


Y
VP16
TAD
X
LexA
DBD
LexA op
Tet R
Tet R
Tet op
HIS 3
No Growth
Y
VP16
TAD
X
LexA
DBD
Tet R
HIS 3
Tet op
Growth